Evaluating Computer-Supported Social Learning:

An Integrative Study of the Development of Soft Skills

in a University-Based Mediation Training Programme

Doctoral Dissertation

submitted in fulfilment of the requirements for the degree

„Doktor der Philosophie“ (Dr. phil. / Ph.D.)

to the

Universität Hamburg / University of Hamburg, Germany,

Fakultät für Erziehungswissenschaft, Psychologie und Bewegungswissenschaft

(Faculty of Education, Psychology and Human Movement),

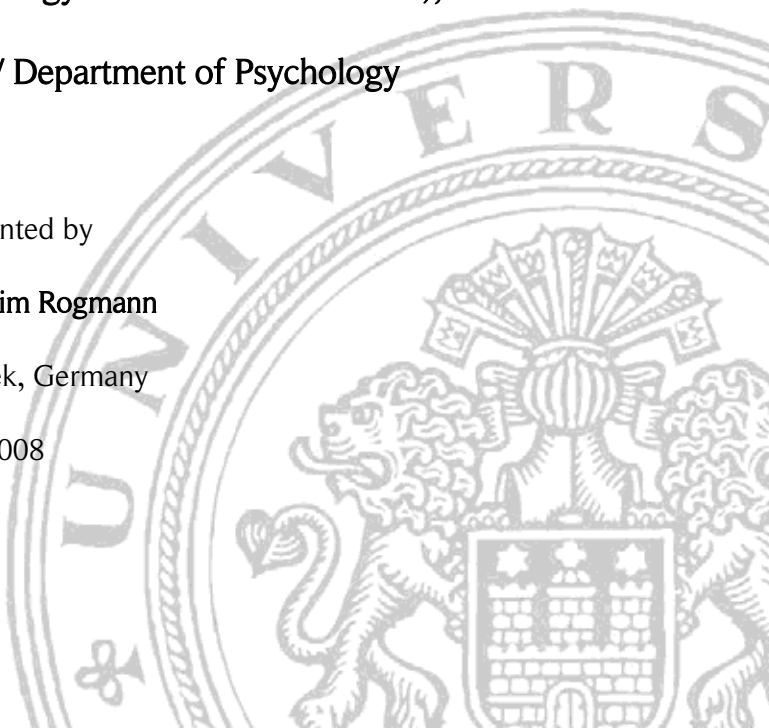
Fachbereich Psychologie / Department of Psychology

presented by

Jens Joachim Rogmann

Halstenbek, Germany

2008



In accordance with §§8 and 9 of the *Promotionsordnung* (Doctoral Degree Regulations) of August 20, 2003, this document had been submitted for review and approval by the Department of Psychology's appointed

Promotionsprüfungsausschuss / Doctoral Degree Examination Committee

Prof. Dr. Reinhold Schwab

(Prüfungsausschussvorsitzender / Head of the Examination Committee)

Prof. Dr. R. Alexander Redlich

(Betreuer und Dissertationsgutachter / Advisor and Thesis Reviewer)

Prof. Dr. Friedemann Schulz von Thun

(Dissertationsgutachter / Thesis Reviewer)

Prof. Dr. Matthias Burisch

(Disputationsgutachter / Disputation Examiner)

Prof. Dr. Torsten Meyer
Fb Erziehungswissenschaften, Juniorprofessor für Erziehungswissenschaft
unter besonderer Berücksichtigung der Forschung und Lehre im Bereich Multimedia

(Disputationsgutachter / Disputation Examiner)

Disputationsdatum/Date of Disputation: 31-03-2008

Evaluating Computer-Supported Social Learning:

An Integrative Study of the Development of Soft Skills

in a University-Based Mediation Training Programme

Jens J. Rogmann

Acknowledgements

As happy as I am now, just having finished the concluding sentences of this dissertation, I also feel grateful to the many without whom this thesis would have never seen the light of day.

First and foremost, I would like to thank Prof. Dr. R. Alexander Redlich for giving me the opportunity to evaluate and contribute to the Department of Psychology's *PROCON* curriculum. He has supported and encouraged and educated me in many more ways than I could possibly state here. Yet, if, humbly, I had to single out one of these many ways for all of which I am most grateful, it would be the opportunity to repeatedly experience how he, based on compassionate and humane principles, proactively and frankly approached and re-conceptualized difficult situations while involving his counterparts (or, if put in a more spiritual term, his neighbours) in an honest and trustworthy dialogue. It has taught me again that, and the end of the day, what really counts is *Menschlichkeit*.

In a similar way, I owe much gratitude to my family for their love and understanding which have proved so essential over all these years. Katja, my beloved wife, and my mother Helga have both encouraged me and have provided unstinting, generous support in both word and deed. Thank you so much for your empathy and all tangible help (which does mean to include child-minding, financial contributions, and also my gracious brother Jan's seemingly unwearied acceptance of extra duties in our family business.) And, Lina, my beloved daughter, thank you for the joyfulness and happiness you have brought to our lives.

Furthermore, I would like to express my thankfulness to my colleague and friend Sascha Kilburg for most enlightening discussions and his help in advancing CSSL theory, coursewares and practice, to Alexandra Bielecke for all her kind contributions and skilful assistance in running the *PROCON* curriculum, and to the other members of the *Beratung und Training* work group for their supportive supervision and making me feel part of a team. I am also grateful to the members of my Doctoral Degree Examination Committee, Prof. Dr. Friedemann Schulz von Thun, Prof. Dr. Matthias Burisch, Prof. Dr. Torsten Meyer, and Prof. Dr. Reinhold Schwab from the University of Hamburg whom I had the honour to learn from and work with in various fields, ways and contexts.

And, finally, I would like to take this opportunity to thank all participating students and peer-tutors for their time, thoughts, and encouragement. Without your remarkable interest in the subject matter, your cooperativeness in contributing, and your honest feed-backs, this project would lack its empirical grounding.

Abstract

There is a growing recognition that graduates from institutions of higher education need to be equipped with higher levels of certified social skills. Recent innovations in advanced learning technologies have provided opportunities for enhancing traditional delivery modes of planned social skill development. While current mainstream research and practice in this area explores approaches to computer-supported collaborative learning (CSCL), an additional framework is put forward in this study.

Computer Supported Social Learning (CSSL) draws on Social Learning Theory and various models for social competence, its development and assessment. CSSL essentially focuses on individual learners and aims at scaffolding situated conceptual knowledge to improve socio-communicative competence. For this purpose, computer technologies provide individual learners with video-based scenes and interactive tasks pertaining to complex and critical inter-personal situations.

In this study, the impact of the introduction and use of tailored CSSL-based coursewares for mediation was evaluated in a university-based setting. Blending traditionally delivered simulation training with preparational use of the coursewares was expected to positively effect learning and resulting levels of socio-communicative competence required by mediating third parties in group-based conflict-resolution.

Across 195 participating students from four successive yearly cohorts, courseware use was found to be associated with superior post-curricular situational judgement. No consistent significant associations were found between CSSL courseware use and post-test levels of conceptual knowledge, interest in subject matter, or self-efficacy.

Potential problems of the statistical, theoretical, and internal validity as well as the generality of the findings are discussed, and ideas for theory, future research and practice are explored.

Table of Contents

… (cont'd)

… (cont'd)

List of Tables

List of Figures

Evaluating Computer-Supported Social Learning: An Integrative Study of the Development of Soft Skills in a University-Based Mediation Training Programme

1. Introduction

It is a unique and original feature of this study to attempt to systematically evaluate the conditions of computer-supported mediation training in a progressive higher learning setting. In this introductory chapter, the target problems to be studied are outlined. Given that mediation training can be understood as a special form of interpersonal or socio-communicative skills training, this study seeks to explore the question how new learning technologies in the form of preparatory video-based, interactive media can support learning.

In the following, interpersonal skills as needed for third-party roles in dyadic and group conflict are briefly reviewed, and a structure of the skills is formulated. Subsequently, the present state of research dealing with the use of instructional and learning technologies to support the development of the skills needed is summarized.

It is contended that - in view of the scarcity of research investigating the effects of blending traditional with technology-based soft skills training delivery formats -, it seems a worthwhile effort to put forward and test a model for evaluating computer-supported soft skills training.

1.1 Mediation Training as Soft Skills Development in Higher Education

Today, "soft skills" or, more specifically, interpersonal or socio-communicative competencies are recognized as being of high relevance in many applied contexts such as psychotherapy (e.g. Beutler, Machado, & Neufeldt, 1994) or management (e.g. Penley, Alexander, Jernigan, & Henwood, 1991; Cameron & Whetten, 1983). Social competent behaviour is also regarded as a keystone for successful interpersonal conflict management (Bowling & Hoffman, 2000, 1989; Ogilvie & Carsky, 2002; Schreier, 2002; The Test Design Project, 1995). It is therefore not surprising that, researchers, representatives of industry and commerce, and educational policymakers alike have repeatedly argued for the implementation of competency-based training and assessment into existing curricula in higher and continuing education settings (e.g. Chur, 2004; DIHK, 2004; Eurydice, 2002; Agiunis & Kraiger 1997).

This need has been described as a shortage of key skills, and, more specifically, as a mismatch between required and actual communication and cooperation skills of higher education graduates (Schaeper & Briedis, 2004; García-Aracil, Mora, & Vila, 2003). This common, transnational problem is also reflected in the European higher education area's adoption of a system of convergence in higher education to be implemented by 2010, often referred to as the Bologna process. To promote graduates' employability, accreditation provisions for new curricula in higher education require a significant percentage of learning credits to be earned through key and soft skills training (cf. e.g. FIBAA, 2000; ZevA, 2000).

This poses a number of challenges to staff, curriculum managers, and policymakers in institutions of higher education (Redlich & Rogmann, 2007). It is a *quantitative* challenge in that, on the one hand, a considerable number of additional courses focusing on key qualifications will have to be added to extant curricula. On the other hand, qualified staff is needed for implementation meaning that either current staff will have to be developed or new teaching staff is to be found. Moreover, it is also a *qualitative* challenge in that the existing and modules and courses of study will have to be developed in terms of content to concurrently provide opportunities for development of key skills. And, finally, it is also a challenge in terms of *didactics and methods*. For centuries, institutions of higher education have focused on refining knowledge development, knowledge acquisition and the passing-on of knowledge. By contrast, fostering – and, subsequently, testing (McClelland, 1973) and credentialing – skills and competencies rather than knowledge has tended to be sidelined in most disciplines, and it

is probably better explored in vocational and professional educational contexts such as training and assessment for management, sales, or customer service.

Methods employed in social skills training.

Chiefly conducted in "classroom" or "face-to-face" settings (Henninger, Hörfurter, & Mandl, 2001), social and communication skills training programmes employ interactive exercises and group discussions (Bedell & Lennox, 1997; Bishop & Taylor, 1992; Fittkau & Schulz von Thun, 1994; Günther & Sperber, 1993; Schulz von Thun, 1984, 2004a). Role-play simulations are also regularly made use of, intended to promote valid behavioural and emotional development and a foundation on which learning through modelling and observation as well as reflection can be built (e.g. Davis & Corley, 1996; Mock, 1997; Redlich & Elling, 2000). Often, simulations are video- and/or audio-recorded to allow for a more in-depth instructor- and/or peer-based feed-back, collective reflection of practice (cf. Schön, 1983), or micro-training (Daniels, Rigazio DiGilio, & Ivey, 1997). Methods of in-class communication training also include other experiential learning exercises (D. A. Kolb, 1984) and practice techniques designed to promote self-awareness and self-enquiry (Bedell & Lennox, 1997). The latter include, for example, approaches based on the "Inner Team" metaphor (Schulz von Thun & Bossemeyer, 1993; Ulrichs, 2004; cf. also Redlich, 2004b, for the mediation context), or the use of self-evaluative questionnaires and journal writing (LeBaron Duryea & Robinson, 1994).

Mediation training as "soft skills" development.

One special form of social competence or "soft-skills" development for applied professions such as Psychology or Law is mediation training (cf. Falk, 2000). Around the globe, mediation and Alternative Dispute Resolution (ADR) training programmes are offered by a growing number of higher education institutions (Botes, 2004; Polkinghorn & Chenall, 2000, Warters, 2000) as well as in private agencies or corporate training institutes.

Regularly, existing programmes include sophisticated communication training (Raider, Coleman, & Gerson, 2000; Redlich & Elling, 2000; Schneider, 2000), methods intended to enhance emotional and self development (Prokop-Zischka, 2000; Reilly, 2005; Schreier, 2002; Shearouse, 2003; Stains, 2003) and opportunities for self-reflection (Marsick & Sauquet, 2000; Picard, 2003). In general, therefore, mediation training curricula draw on the same instructional

techniques and, possibly, learning paradigms as employed in other fields of human resource and soft skill development (some of which are outlined above).

Some justifiably call for mediation training that aims more directly at developing the trainees' personal skills in handling emotions as well as their emotional self-awareness and self-regulation skills (Schreier, 2002). However, in a precursory study of the mediation training curriculum evaluated here, Langhorst (2005) suggests that mediation training curricula may have the potential to further the personal social skills of curriculum participants despite their different focus on the role of neutral third parties, for instance in the form of enhanced individual negotiation skills and increased self-efficacy in individual dispute handling.

The need for evaluative research.

Albeit today there is an ever-growing body of research and practice literature dealing with mediation, its models and processes, the role of third parties in conflict management, mediation skills, and even the training of neutrals, few researchers have yet attempted to evaluate learning progress and outcome in mediation training programmes.

In the on-going process of professionalization in mediation (Maiwald, 2004; Picard, 1994), professional societies as well as juridical and governmental bodies attempt to channel this very process by issuing training guidelines and accreditation standards (Wassner, 2002). In good accordance with their objective to provide a basis for quality in mediation (SPIDR Board of Directors' Commission on Qualifications, 1995), many of these guidelines address the evaluation of trainee progress and trainer responsibilities (cf. e.g. Pou, 2002). Here, the solicitation of evaluative comments from trainers or trainees, the assessment of participants, and the use of training evaluation forms are called for (e.g. CAADRS Center for Analysis of Alternative Dispute Resolution Systems, 2004; Conflict Management in Higher Education Resource Center, 2004) or even mandated in court programmes. Interestingly, in contrast to these pleas, a comprehensive mediation training program evaluation has not yet been an objective in applied psychological, educational, or social research. In 2000, Raider, Coleman and Gerson found that

> "although there is an extensive theoretical and empirical literature on the nature of conflict and the processes of negotiation and mediation as applied in diplomacy, business, and labor relations, there is very little systematic research on the pedagogy of conflict resolution or on the models and methods used to teach these skills to adult or student learners" (p. 499).

Since, a growing body of research has further explored the field of in-school conflict resolution education (Deutsch, 2000; Jones, 2004), manifested primarily in peer mediation programmes (e.g. Burrell, Zirbel, and Allen, 2003) and "life skills" programmes aimed at developing student's emotional, cognitive, and behavioural competencies thought to be required to deal with personal conflict (e.g. Greenberg et al., 2000). Overall, however, there still is extremely little 'developmental' research concerned with helping to shape effective educational and training programs in the conflict resolution arena (Deutsch, 2000). In addition to difficulties evaluation researchers typically come across in higher education settings (e.g. cf. Bülow-Schramm, 1995; Lohnert & Rolfes, 1998; McEvoy & Buller, 1990; and, discussing specifics of e-learning evaluation in higher education, Schwarz, 2001), a number of additional reasons could account for this phenomenon.

No evaluation tradition.

Firstly, in the relatively young field of conflict resolution, mediation program review processes in higher education are usually an "internal matter" and curricula, course content, student performance standards, or program structures are usually not subjected to evaluation (Botes, 2004).

Vocational assessment of practitioners.

Secondly, as it is a common notion in the field that training is only one of a multiplicity of paths to acquire mediation skills, much emphasis has been put on defining mediator competencies and developing methods to assess these (Friedman & Silberman, 1993; Herrman, Hollett, Gale, & Foster, 2001, 2002; Honeyman, 1990, 1993; Honoroff, Matz, & O'Connor, 1990; LeBaron Duryea, 1994; Matz, 1993; McEwen, 1993; Pou, 2002; SPIDR Board of Directors' Commission on Qualifications, 1995; The Test Design Project, 1995; Wassner, 2002) rather than evaluating the acquisition of these skills in training programmes. Moreover, despite all attempts to pinpoint knowledge, skills, and other attributes (commonly referred to as "KSAOs") needed for successful mediation, recent articles suggest that it remains to be debated whether and which KSAOs should or could be tested or, indeed, if they are prerequisite for the employment of mediators at all (Honeyman, Go, & Kelly, 2004).

Foci of mediation research.

Thirdly, much evaluative research in the field has focused on determining the efficacy and effectiveness of mediation programmes, while theory-driven dispute resolution research

has accentuated experimental research strategies. In doing so, both have generally contributed little to either practice or mediation training (McEwen, 1999).

Funding of mediation research.

And, fourthly, while there may be an interested consumer audience for research aimed at evaluating programmes aimed at qualifying neutrals, evaluation research with the objective to go beyond mere efficacy control typically requires additional funding and thus, the need to accept additional evaluation stakeholders.

In sum, there is a dearth of educational evaluative research, however. Fresh contributions to the field aimed at evaluating training programs for mediation that can help enhancing existing programs are highly warranted.

DIGEST 1.1

It is a challenging undertaking to introduce competency-based trainings for key qualifications into the curricula offered at higher education institutions across Europe. These training courses employ specific methods that aim to reduce the mismatch found between the level of social skills required of graduates entering professional life and their actual skill levels. One special form of social skills training is mediation training. For various reasons, there is a dearth of educational evaluative research in this area; innovative contributions are therefore highly warranted.

1.2 Social Skills Training: 'Face-to-face' and 'Blended' Delivery

The past decades have seen an unanticipated evolution of learning and instruction by means of electronic media, such as audio and, later, video tapes, -disks, or –broadcasts, computer-based applications with an increasing degree of interactivity, the use of the Internet and other forms of digital media. The utilization of these media for learning purposes have greatly affected (a) the ways learning content is distributed or delivered, (b) the potentials of learner control, adaptive tutoring, and collaboration in learning, and (c) teaching activities. It will be explored below, in what ways and to what extent technology may complement 'traditional' social skills training methods as outlined above.

In the course of technology advancement, a vast variety of expressions have been coined to characterise the phenomenon that traditional learning activities and learning management are today supplemented by – or even replaced with – technological means. Some of these expressions are "computer-" or, more broadly, "technology-assisted (or –aided, -enhanced, or -mediated) learning", "e-learning", "computer-" or "web-based learning or "training", "hybrid" or "blended learning", "online", distance" and "multimedia learning", "hypermedia-" or, shorter, "hyper-learning", "open learning system", "flexible" and/or "distance learning", etc.[1] The same holds true for measurement and assessment processes in education and personnel selection, with the "learning" or "training" part of the expressions listed above being replaced by "assessment", "testing" "evaluation", or "measurement".

As Goldman-Segall and Maxwell (2003) explain, the functional roles technology can have in learning are manifold. Technology can serve as an information provision system, as a subject and curriculum area in itself, as a communications medium, as a "thinking tool", as an experiential environment, as a developmental scaffold, or as a discourse and perspectivity toolkit. However, they also note that research often has "the tendency … to use an instrumentalist and instructionist approach" (Goldman-Segall & Maxwell, 2003, p. 397) with little regard for social dimensions of learning. As a result, learning technologies have often been viewed as a possible substitutes for "traditional" learning activities (often termed "face-to-face-" or "classroom" learning), and, consequently, researched as such: Numerous

[1] It seems to depend on tradition, fashion (Northrup, 2002) and progress of technology (Reinmann-Rothmeier, 2003a) yet which expression is being used and what exactly is meant by it. At this point, the term "e-learning" will be used in the following. Subsequently, for the purposes of this study, more detailed definitions will be given below.

comparison studies and meta-analyses (e.g. Azevedo & Bernard, 1995; Burns & Bozeman, 1981; Cohen & Dacanay, 1992; Fletcher, 1990; Fletcher-Flinn & Gravatt, 1995; Hannafin, 1985; Joliceur & Berger, 1986; Kulik & Kulik, 1989; 1991; Kulik, Kulik & Cohen, 1980; Lou, Abrami, & d'Apollonia, 2001; Niemiec & Walberg, 1987; Olson & Wisher, 2002; Roblyer, Casting, & King, 1988; Ryan, 1991; Russell, 1999; Shavelson, Webb, & Hotta, 1987; Sitzmann, Wisher, Stewart, & Kraiger, 2004; Sitzmann, Kraiger, Stewart, & Wisher, 2006) have been conducted with the aim to determine whether technology-based learning could – under which circumstances – substitute face-to-face instruction (cf. e.g. Schulmeister, 2002a, for a critical overview). Whether new media actually make a significant difference or not, remains, however, a subject of controversial debate (cf. e.g. Phipps & Merisotis, 1999; Ramage, 2002; Schulmeister, 2002a; Welsh, Wanberg, Brown, & Simmering, 2003). Perhaps it is not even possible to create experimental training conditions that are identical in every single respect but form of delivery (i.e. face-to-face vs. computer- or web-based), and comparative studies may therefore inevitably be flawed (R. E. Clark, 1994).

Yet, recent contributions to the field lend support to the notion that new media should rather be viewed as *supplements* or *complements* than as *substitutes* to traditionally known systems of learning and instruction.

Potential advantages of "blending".

Zenger and Uehlein (2001) caution against letting "…the disciplines of instructor-led training and e-learning parallel each other when there can be huge gains through integration" (p. 57). While, as Carman (2002) notes, e-learning "[…] industry consensus continues to point to the use of multiple modalities for learning" (p. 1), the view that "[…] people perform better when they have a mix of modalities and methods of learning" (ibid.) is generally known among practitioners and researchers alike (e.g. D. Clark, 2003; and Hasebrook, 1997) to be a naïve theory assuming that there is a simple addition of effects. In their recent review of the available blended learning literature, Graham, Allen, and Ure (2003) identified three major recurring themes often referred to as potential benefits to blended or "hybrid" delivery formats, namely, (1) increased pedagogical effectiveness, (2) improved accessibility and flexibility, and (3) cost effectiveness. In these areas, traditional instructional methods and e-learning may have different strengths that could complement each other.

For example, the inclusion of technology based delivery may save learner travel time and cost, may allow for a higher degree of flexibility, an individualised learner progress, and,

by means of simulation, less expensive experiential learning in a realistic albeit safe environment (cf. e.g. Bainbridge, 1995; D. Clark, 2003; Meifert & Piehl, 2000; Sauter, Sauter, & Bender, 2004; Voci & Young, 2001; Zenger & Uehlein, 2001). In particular, this may be the case where technologies allowing for asynchronous formats of communication reduce the monetary and non-monetary outlay of individual learners (Kerres & de Witt, 2003). Results further suggest that lecture-based and computer-based learning may target at different skills (Williams, Aubin, Harkin, and Cottrell, 2001; cit. in Berger, 2004). The incorporation of hypermedia-based formats might also be favoured by instructors due to its comparative pedagogical richness, the learner's accessibility to knowledge and the ease with which learning materials can be revised and updated (cf. Osguthorpe & Graham, 2003). And, vice-versa, fully replacing face-to-face delivery with equated technology-based solutions may be a costly endeavour (Shaw & Young, 2003). Trying to realize high levels of interactivity and feedback in technology-based solutions lead to a significant increases in the instructors' workloads (Kearsley, 2000).

Social skill development may require situated, direct interaction.

Secondly, a certain amount of direct interaction between instructors and learners and among learners often seems wanted by learners (Dzubian, Hartman, & Moskal, 2004) as well as indispensable to ensure programme quality (Su, Bonk, Magjuka, Liu, & Lee, 2005) and to retain possibilities for an assessment of the individual learner's potentials and the provision of developmental feedback (Farrell, 2000). Moreover, face-to-face interaction may be required to foster a positive attitude of learners towards technology-based learning (Kurtz, Sagee, & Getz-Lengerman, 2003). While technology may be used as an environmental "scaffold" across different teaching contexts (McLoughlin, 2004), in the area of social learning, "[…] scepticism abounds regarding e-learning's capacity to deliver powerful soft-skill … development" (Zenger & Uehlein, 2001, p. 58; cf. also Farrell, 2000). There are several reasons for this.

Social perspectives on learning suggest that "… a learner will always be subjected to influences from the social and cultural setting in which the learning occurs, which will also define at least partly the learning outcomes" (Mayes & de Freitas, 2004, p. 9). For instance, social skills development may require opportunities for cultural exchange (Vygotsky, 1978) and collaborative experiential testing (cf. D. A. Kolb, 1984). For example, the shared experiences induced by in-class role-plays and experiential exercises often serve as a reference for the more abstract vocabulary, the scientific terms and the theoretical concepts the learners acquaint

themselves with. It can hardly be imagined how a common understanding and a shared meaning of the taxonomy should come about without social interchange.

Moreover, sustainable and effective development of social and communicative skills may necessitate coherence in that (a) learners preserve and develop (selective) authenticity in their behaviour and (b) that their behaviour meets situational prerequisites (Schulz von Thun, 1984, 1998, 2004a). Both aspects profit from public accountability and personalised feed-back in identifiable, credible, and trustful relationships, and direct and rich forms of interaction, possibly even embedded in an active community of practice (cf. Lave & Wenger, 1991). While all of these features chiefly can genuinely be cultivated in face-to-face settings, contrivance in both solitary and collaborative technology-based learning seems seldom realistic or practical.

Zenger and Uehlein (2001) also point out that face-to-face, group-based training is culturally anchored and a well-tried method that draws on human preferences and socialised school experiences. It may also provide for motivating experiences, complex and unintentional interactional learning in ways computer-based learning in itself cannot offer.

Especially in programmes aimed at developing more complex social skills and competencies, positive comparative effects are assumed for blending (cf. Dittler, 2002; Evans, Sparkes, Jordan, Jones, Chase, and Curtis, 2001; Jenkins Henninger, Hörfurter & Mandl, 2003). Indeed, in their recent review of technology-assisted learning formats for management skills, Arbaugh and Stelzer (2003) concluded that initial comparative studies indicate "[…] hybrid formats compare favourably" to both straight classroom and online modes of learning (p. 19). However, these assumptions are based on blended learning outcome studies employing technologies which could be regarded as outdated today (e.g. McNeil & Nelson, 1991) or are generalizations of studies investigating the acquisition of knowledge or simpler skills such as software application (e.g. NETg/Thomson, 2002). In summary, both the utilisation of software for the development of professional soft skills and e-learning research in this area are presently still in its infancy. The lack of more complex and recent blended learning research can be attributed to a variety of factors outlined below.

"Blending" as an emerging trend.

Sauter, Sauter, and Bender (2004) as well as Reinmann-Rothmeier (2003b) view the combination with face-to-face learning as a relatively recent development in the practice of e-learning. Likewise, in a recent review, Welsh et al. (2003) speak of a "trend toward

increasingly blended solutions that use a combination of asynchronous, synchronous and classroom experiences" (p. 255). This perspective can be understood as originating from a distance education approach or an e-learning or point-of-view, spawned by the – largely unfulfilled (Tergan, 2004) – hopes to create compelling e-learning substitutes for face-to-face learning. However, one could also adopt the position that traditional or "classroom" learning had long been the only form of delivery – even, as D. Clark (2003) notes, "[…] with no reflection at all on the issue of media selection […]" (p. 9). Then, suddenly, every supplement to face-to-face learning by way of inclusion of major technological innovations (in their chronological order: phonetic alphabets, papyrus and paper, printing/books, broadcast media, mass media storage devices, personal computer, networks; cf. D. Clark, 2003) to could be termed a "blend" in its time. In this sense, "blended learning"[2] has always been around, irrespective of learning technology use, and today signifies the efforts of suppliers of "traditional" learning (such as universities and corporate training institutions) to incorporate and experiment with elements of e-learning.

It was predicted that, between 1997 and 2000, companies would reduce classroom-based training by nearly 20%, replacing it with technology-based forms of training (Bassi & Van Buren, 1999, cited in Kraiger, 2003). However, experts estimate that, in 1999, only 2% of training was web-based, and that 75 percent of that was in information technology (Moran & Allerton, 2002). Recent surveys suggest that, nowadays, possibly more than half of all corporations offer up to 40% of their human resource development courses in formats supporting classroom-based learning with technology-based delivery (Bonk, Kim, & Zeng, 2006; The eLearning Guild, 2003).

When compared to corporate settings, the application and utilization of blending "traditional" and "e-learning" in adult training programmes seems less progressed in the higher education area. A recent study indicates that the introduction and use of technologies required for blended learning (such as web-based courses and training CD-ROMs) is much further advanced in business, healthcare, and governmental institutions when compared to university settings (Gemeinhardt, 2002). There may be various reasons for this. In their international survey of the use of information and communications technologies (ICT) in the higher

[2] However, if defined either as a vast or an "ideal" variety of combinations of two (or more) more or less distinct methods of training, the term "blended learning" is prone to become a buzzword unredeemable for all attempts to derive any substantial amount of meaning from it . For the purposes of this study, a careful delineation of the terminology in use seems therefore warranted and will be undertaken in Chapter 2.

education sector, Collis & van der Wende (2002) conclude that both institutional programmes and their instructors feel, by large, not fundamentally affected by educational technologies. Moreover, traditional forms of delivery (e.g. lecture) are used predominantly and are highly valued. Despite the huge amounts that have been invested into the whole range of ICT in institutions of higher education, blends with ICT are only gradually introduced, often through personal motivation and "bottom up" initiatives of single instructors and regularly without particular reward (Collis & van der Wende, 2002). In his critical analysis, Selwyn (2007) points out that this situation may also be due to a severely limited, linear and feature-oriented image of advanced learning technologies that has been produced by the various higher education stakeholders dealing with the introduction and use of ICT by following their vested interests and idiosyncratic rationales.

If it is true that the rate and extent of ICT implementation is lower in university settings (Gemeinhardt, 2002), then educational researchers are just beginning to have access to samples and settings of convenience as the university setting continues to provide accepted and manageable environments for studies of learning and instruction. Furthermore, technology has advanced considerably in the past five years with particular ramifications for the soft-skills development domain. Until the late 1990s, a major problem of soft skills training software was its limited interactivity. Despite an ample supply of videotapes for socio-communicative learning purposes in general and mediation training in particular, interactive instructional software was hardly available (Gentry, 1992). Due to lacking bandwidths and limited computer processing speeds it seemed impossible to mimic or present human interaction with a sufficient degree of realism (Barron, 1998). Nowadays, however, the general technological infrastructure in developed countries and, more specifically, in institutes and environments of higher education in particular does usually permit the utilization of interactive, video-based courseware and internet applications.

Meanwhile, however, the interests of many educational technology researchers seem to have shifted away from computer- or net-based hypermedia and asynchronous training applications towards synchronous, real-time and collaborative tools (Koschmann, 1996, 2001; Schulmeister, 2004).

Preliminary Research Outcome.

Herein lie possible reasons why large-scale investigations into blended learning have so far been carried out predominantly by ICT vendors and corporate enterprises.

In the Thomson Job Impact Study (NETg/Thomson, 2002, 2003), the effectiveness of blended learning formats for corporate software and fundamental business skills training was measured against the e-training-only and no-training options. The study involved 200 employees at all organizational levels across a wide range of industries. At post-test, the average accuracy of skill performance in the blended groups was much better than the no-training group and somewhat better than the e-learning-only group, indicating that a blended learning system can also be superior to an e-learning-only delivery format.

Bersin and Associates (2003) documented corporate cases that have effectively used blended learning to provide a large return on investment (ROI). In their research project, they conducted interviews with 17 large corporations who had developed major blended learning programmes for customer education, product and sales training, accounting practices, and manufacturing processes. A business-critical problem, a large, dispersed audience, and a short delivery time are viewed as vantage points for successful and cost effective blended learning solutions (cf. Graham et al., 2003).

Among the university-based research papers that address blended learning is Goldberg's (1997) comparison study with 70 students taking a third-year course in computer science. The study yielded anecdotal evidence that blended learning may be advantageous in terms of academic results and in terms of student acceptance when compared to traditional face-to-face only or web-based only delivery formats. Another study (Dean, Stahl, Sylwester, & Peat, 2001) found that the self-reported learning gains of executive physicians enrolled in a continuing education MBA programme were significantly higher in blended systems of delivery as compared to traditional face-to-face instruction. By contrast, in a controlled quasi-experimental study comparing traditional versus blended delivery for software skills and language training in a vocational setting, Remdisch, Heimbeck, and Kolvenbach (2000) were unable to find significant between-group differences in performance-based tests after training. Similarly, in a university-based evaluation of blended learning modules for basic communication skills, Bildat (2003) reported that no significant differences could be found between traditional classroom delivery and a combination of web-based preparation with face-to-face training.

However, evidence from various case studies in the communication skills training domain indicate that replacing face-to-face learning with video-based practice materials regularly increases levels of self-awareness and self-efficacy (Cauble & Thurston, 2000; Evans

et al., 2001; Fleetwood, Vaught, Feldman, Gracely, Kassutto, & Novack, 2000; F. C. B. Hansen, Resnick, & Galea, 2002; Koerfer, Thomas, Obliers, & Köhle, 1999; Resnick, 1998; Wiecha, Gramling, Joachim, & Vanderschmidt, 2003).

In spite of these promising findings, scholars and practitioners alike lament the generally unsatisfactory research situation in the blended learning field. Recently, Dewar & Whittington (2004) noted that "while there is a great deal of literature defining blended learning and making suggestions for how to implement it, there's less dealing with the actual effectiveness of blended learning. […] The biggest challenge is finding studies that specifically address blended learning, as opposed to the use of technology alone" (p. 5). More original studies are needed to evaluate the effects and efficacies of blended learning solutions.

For these reasons, it has become a major objective of this study to develop and implement a practice-oriented evaluation system that allowed for studying the effects blending existing socio-communicative skills training modules with computer-based preparation modules. A theoretical framework for evaluating computer-supported social skills training will be developed and put forward below, which serves as a basis for the exploration of technology-supported development of socio-communicative competence.

DIGEST 1.2

Today, traditionally known systems of learning and instruction for the development of social skills are being supplemented by technological means. In view of the scarcity of research investigating the effects of blending traditional soft skills training delivery formats with technology-based training, it seems a worthwhile effort to attempt to evaluate the potential advantages, the conditions and outcomes of hybrid delivery in a progressive higher learning setting.

2. Theoretical Background

Summarizing the above, little is known about the question, why and in what ways technology enhance can training programs for soft skills, or, more specifically, mediation training. This study seeks to make a fresh contribution to the field through (a) reviewing implications of current theoretical approaches to technology-assisted learning in the soft skills domain, and (b) by presenting a practice-oriented evaluation system that permits a theory-driven study of the effects brought about by the implementation of a specific hybrid delivery format into a pre-existing mediation training curriculum.

The theoretical background of this study will be detailed in the following. To delineate the scope and objectives of the approach taken in this study, a model for understanding the nature of socio-communicative competence and its assessment will be outlined first.

Subsequently, current theories of learning relevant to understanding the development socio-communicative competence will be examined as to their advisory and explanatory strengths for the subjects under study, namely soft skills training, blended training delivery and instructional evaluation.

This study's underlying integrative framework for socio-communicative competence development and resulting blended learning strategies in this area will be presented in the conclusion of this section.

2.1 The Nature of Socio-Communicative Competence

As Kanning (2002) notes, the terms "social skills" or "social competence" are often used synonymously. Such is the case for related terms, concepts, and constructs such as "soft skills", "emotional intelligence" (Bar-On, 1997; Goleman, 1995; Salovey & Mayer, 1990; Sjöberg, 2001), "social intelligence" (Marlowe, 1986; Sternberg, 1985; Thorndike, 1920), "cultural intelligence" (Earley & Ang, 2003), and, from a more sociological point of view (cf. Opengart, 2005), "emotion work" or "emotional labour" (Hochschild, 1983). Moreover, depending on the research tradition of the respective researcher and the intended use of a term largely defines its connotation and its evaluative frame of reference. As Weinert points out, "the many implicit (in word use) and explicit (in theoretical frames of reference) definitions of competence are so heterogeneous that only a small, vague conceptual core remains" (1999, p. 26). Indeed, the term "competence" itself is accompanied with a similar level of fuzziness (e.g. Draganidis & Mentzas, 2006; Grzeda, 2005; Westera, 2001). Consequently, for the purposes of this study, a working definition is warranted aspects of which will be outlined in the following.

Social competence as Socio-communicative competence.

Holsbrink-Engels (1998), with reference to Ellis and Whittington (1981), distinguished between three distinct research perspectives or traditions for socio-communicative competence, namely (1) developmental approaches focusing on the development of socially skilled behaviour in childhood and adolescence; (2) remedial approaches concerned with adaptive behaviour failures and intervention therein; and (3) a specialized approaches attending to socio-communicative skills in professional interactions. It is the latter perspective this thesis will focus on, with special reference to skills required for the management of interpersonal conflict as a third party.

In general, (behavioural) competencies can be broadly conceptualised as prerequisites commonly attributed to an individual person which are needed to effectively deal with demands manifested in and across a number of specific situations or "task domains" (Weinert, 1999, p. 5). As a preliminary working definition, therefore, social competence can be understood to be a sub-category that includes (1) individual prerequisites (2) needed to effectively interact with other people in (3) specific types of interpersonal situations. Inter-personal interaction here is conceptualized as including both dyadic interaction between an individual and another individual as well as interaction between an individual and a group of

other individuals (cf. Holsbrink-Engels, 1998). Therefore, in the following, the broader term "social communication" will be used to denote the interactive processes. Accordingly, the expression "socio-communicative competence" rather than "social competence" will be used. The three inter-related aspects of socio-communicative competence will be discussed in the following.

Multidimensionality in individual pre-requisites.

Socio-communicative competence is seen as a multidimensional construct, in correspondence with most publications on the subject (cf. Kanning, 2002; Le Deist & Winterton, 2005). The working definition set out above refers to socio-communicative competence as being an array of characteristics attributed to an individual. Formally, this array refers to subject-specific, unremittent, dispositional traits that, in a medium term, remain stable across the same type of situations (Euler & Reemtsma-Theis, 1999).

Implicitly, this concept endorses a view of people of as "agents". From an agentic perspective, individual communicants (or "protagonists") are seen as having intentions and the capabilities of forethought, self-reactiveness, and self-reflectiveness (Bandura, 1999, 2001), or, in a more general term, self-regulative capability (Karoly, 1993; Ryan & Deci, 2000b; Zimmerman, 1989).

The myriad of prerequisites that could possibly be named do include knowledge, cognitive abilities, and psychomotor skills, much in line with traditional taxonomies of educational activities (e.g. Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). As for other competencies, socio-communicative competence can be thought to require a specific set of knowledge (i.e. declarative or propositional information, conceptual understanding, and tacit or procedural knowledge: "know that/what" and "know how"), attitudes (e.g. self-efficacy and goal content; cf. Vansteenkiste and colleagues, 2006, 2004), self-determination and skills (i.e. perceptual, response-selection, psychomotor, and problem-solving skills, cf. Proctor & Dutta, 1995).

However, as Euler (2004) points out, socio-communicative competence may differ from other areas of competence in that it is value-laden and dependent on cultural norms. The concept of socio-communicative competence thus relates to attitudes and other characteristics of the individual (e.g. emotional skills, values) which are generally not subsumed under the knowledge and skills headings. Indeed, in their typological model of competence, Le Deist and

Winterton (2005) see 'social competence' as one of four dimensions of competence in general which is related to personal functioning. Social competence can thus be conceptualized as interrelated with the other three types of competence, namely (a) 'cognitive competence' or occupation-related conceptual understanding or *knowledge*, (b) 'functional competence' or task-, function-, and role-related operational *skills*, and (c) 'meta-competence', a set of individual conceptual skills and behaviours facilitating personal effectiveness and the acquisition of new competencies. However, Le Deist and Winterton (2005) also point out that "in practice, not only must a person have underlying knowledge, functional skills and appropriate social behaviour in order to bee effective at work, the competences required of an occupation are also invariably described in multi-dimensional terms" (p. 39-40).

In sum, therefore, the individual prerequisites of socio-communicative competence can be thought to comprise knowledge, skills, attitudes, and other personal characteristics, or *KSAO*s as they are often referred to in the human resources domain (Sackett & Laczo, 2003). Any interaction will draw on a variety of *KSAO*s. For example, in order to greet another person, a communicating protagonist needs conceptual knowledge about how to appropriately address the other person (be it a relative, a stranger, a Japanese, or the Queen of England; be it in the presence or absence of others). Depending on the situation, the protagonist would also need the skills to perform the greeting ritual suitable in the given cultural and situational context (be it the shaking of hands with appropriate pressure, the exchange of kisses of the right kind, number, and duration, or the accurately performed bow or curtsy). And the communicant would have to evaluate and balance his emotional, motivational and attitudinal states (be it the despise of welcoming an ill-famed dictator in a diplomatic encounter, the excitement of encountering one’s lover, or the fear of being subjected to critical appraisal by one’s superior).

In an attempt to integrate frequently cited KSAO-inventories of social competence in terms of their verbal meaning, Kanning (2002) put forward a general list of 15 KSAO dimensions in three categories (see Table 1 below). The three categories can be viewed as related to the KSAO categorisation in that they pertain to perception and cognition (the sphere of knowledge), to behaviour (the sphere of skills), and to motivation[3] and emotion (the attitude and other characteristics domains).

Table 1

Dimensions of Social Competence

Perceptive-Cognitive Domain	**Motivational-Emotional Domain**
self-attentiveness (*Selbstaufmerksamkeit*)	• emotional stability (*emotionale Stabilität*)
social perception (*Personenwahrnehmung*)	• prosociability (*Prosozialität*)
perspective taking (*Perspektivenübernahme*)	• value pluralism (*Wertepluralismus*)
locus of control (*Kontrollüberzeugung*)	
decisiveness (*Entscheidungsfreudigkeit*)	
knowledge (*Wissen*)	

Behavioural Domain
extraversion (*Extraversion*)
assertiveness (*Durchsetzungsfähigkeit*)
behavioural flexibility (*Handlungsflexibilität*)
communicative skills (*Kommunikationsfertigkeiten*)
conflict behaviour (*Konfliktverhalten*)
self-control (*Selbststeuerung*)

Note. Table adapted from Kanning (2002), p. 158.

From an instructional perspective, Euler (2004) criticizes that, commonly, the listed prerequisites for social competence remain unrelated to specific behavioural, subject-specific content, and situational demands. For example, rather than seeing "pro-sociability" as an uni-dimensional personal attribute of socio-communicative competence, it may have different meanings in substance or content as to whether a protagonist serves in a charity organisation, as a civil service diplomat, or as the shareholder of a large investment group. Situational demands for "pro-sociability" may also differ as to whether a protagonist is to shake hands with a superior, a client, or an ill-famed dictator. Therefore, the two other aspects of the preliminary definition of social competence mentioned above will need to be considered.

[3] Some authors have argued for an exclusion of motivational dimensions on grounds of keeping the competence construct less complicated. As Winterton, Delamare-Le Deist, and Stringfellow (2005) point out, "motivation is clearly not a part of competence. A person is said to be 'competent' if they have the requisite [KSAOs] …, but whether or not they are motivated is a function of a whole range of external and internal factors" (p. 27). Motivation is paramount in competence development and in assessment performance, however . For the purposes of this study concerned with instruction and measurement, therefore, motivational components will be included separately.

Direct and Effective Interaction.

Moreover, communicative competence in general may be thought to include references to an undefined number of people (e.g. community, society, humankind, etc.) or to "asynchronous" communication entailing potentially considerable temporal lags (e.g. letters, e-mail, weblogs, etc.). However, for the purposes of this thesis, socio-communicative competence will be used to denote direct, synchronous interaction among present individuals,- which may, however, include interaction relayed by means of media (e.g. cell phones, webcams, etc.).

Yet, what does *effective* interaction mean? In his recent review of attempts to define social competence, Kannig (2002) noted that most definitions differ in their accentuation of two different and conflicting set of skills and attitudes, namely, (a) the capacity of individuals to successfully pursue their own goals during the interaction with others, and (b) their capability to adapt to the social conditions of their environment. Interestingly, this division bears many similarities to contemporary conflict management grid models (Blake & Mouton, 1964; Pruitt & Rubin, 1986) and Schulz von Thun's model of dual coherence (*Stimmigkeit*) in interpersonal communication (Schulz von Thun, 1998, p. 306; see also below). It also lends some support to the notion that the management of interpersonal conflict – as a party in conflict, but also as a third party – may well be regarded as central to the concept of socio-communicative competence.

Euler (2004) presents an approach well compatible with the purpose of this thesis. He argues that "social communication", i.e. interactive processes, are the basis and frame of reference for the activation of socio-communicative competence. In other words, interaction presents an individual with requests for communicative action. To handle these requests, the individual requires the KSAOs of social competence (see Figure 1 below).

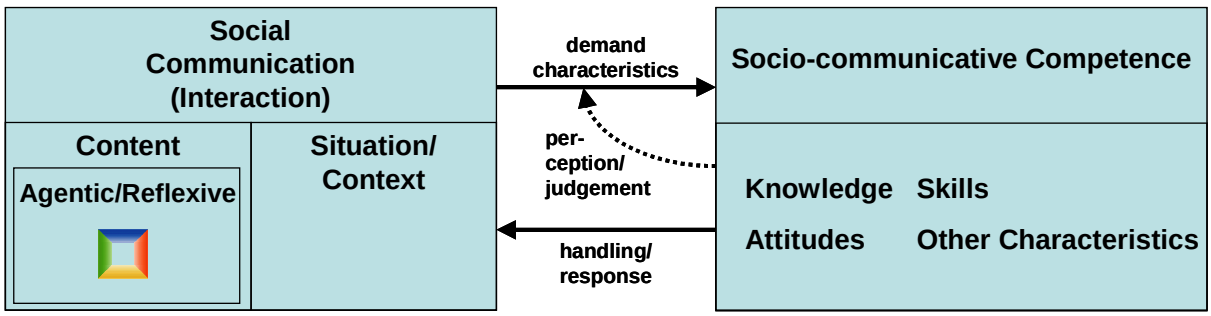


Figure 1. Relationship between Socio-communicative Competence and Social Communication (based on Euler (2004), pp. 13, 44)

Euler (2004) and his colleague Reemtsma-Theis (Euler & Reemtsma-Theis, 1999) also point out that social communication essentially addresses subject-specific contents. They put forward a two-fold model of socio-communicative content which has bearing on the subject under study. In what they call the "agentic" focus (*agentiver Schwerpunkt*) of inter-personal communication, the communicants alternately express themselves, and both view their reciprocal interaction as unobstructed and discord-free. They suggest that the communicative content can be conceptualised in terms of the four aspects (informational content, self-disclosure, appeal, relationship) originally put forward by Schulz von Thun (1981; Schulz von Thun, 2004b).

However, when aspects of communicative content remain ambiguous or covert or if "decoding" processes (perception and/or interpretation) appear narrowed or constricted, the collaborate dynamics of inter-personal interaction may lead to the experience of inconsistencies, dissonance, friction, and/or discord. This is generally in line with the accommodation theory of interpersonal communication (Giles & Street, 1994). A communicant may then enter what Euler (2004) refers to as the "reflective" interactional focus (*reflexiver Schwerpunkt*) in that he or she tries to make sense of this experience and to (re-)interpret the situation. This may happen in both, an open, mutual, meta-communicative exchange, or in internal discourse. In this model, therefore, the handling of the transition between "agentic" and "reflective" modes of social communication is seen as influenced by social competence.

Situational Appraisal.

Interpersonal interaction is always "situated" in that it takes place in a specific context. In the model laid out above, the situational and contextual characteristics shape and substantiate the very demands placed on the communicants. Based on the above discussion, this model can now be elaborated. In order to generate a response, an individual is required to appraise (or re-construe) the often complex situation and make sense of various situational characteristics. For scientific and reflective purposes, one may conceptualise the totality of the interactive situation and its context (which has to be further analysed below) but also its communicative content as a "message" which then can be subjected to further scrutiny, for example, by using the models of Schulz von Thun (1981, 2004b) as Euler (2004) suggests.

However, this conceptualisation may not make explicit enough that processes of social cognition – in similarity to but possibly neurologically distinct from other cognitive processes (Mitchell, Macrae, & Banaji, 2005) – are highly dependent on previous knowledge, extant

cognitive schemas, scripts, the emotional state of the individual as well as attitudinal (pre-)conceptions (see e.g. Ajzen, 2001; Fiske, 1993; Wyer & Srull, 1994a, 1994b; Todorov, 2002). In other words, what aspects of the situational "messages" are being predominantly "received" (i.e., attended to, perceived, understood, and/or processed) largely determines the perceived demand characteristics of the interaction. For example, at a moment in that one individuals feels that that his or her response is being required, he or she will *inter-punct* the interactive process described as social communication (cf. Watzlawick, Bavelas, and Jackson,1967); yet, how does such a "feeling" come about?

It is a feature of professional types of socio-communicative situations that "experts" (or authors of computerised learning environments) typically conceptualise the situational features and, thus, demand characteristics differently from novices or learners (cf. Dawson et al., 1989, Fiske, Lau, & Smith, 1990; Jones & Read, 2005; Thompson, 1990). For the purposes of this study, it will therefore be assumed that the variety and possible permutations of situational features that could possibly influence processing and handling typically exceed the individual's perceptual and/or processing capacities. Therefore, one important aspect of socio-communicative competence is the process of appraisal or judgement of situational features (which itself thought to be guided by an individual's KSAOs). Figure 1 above illustrates these relationships.

Situation and Contextual Conditions.

While it may be argued that any given situation is unique for any of the participants, Euler (2004) suggested that situations may be perceived as similar in their demand characteristics and, thus, categories could be construed that include specific "types" of situations or demand characteristics. In principle, this very idea also underlies longstanding procedures in psychological research and applications, such as Situational Judgement Tests (see chapter 2.2 below), the Critical Incident Technique (Flanagan, 1954), or the Cultural Assimilator (Fiedler, Mitchell, & Triandis, 1971).

To identify the respective demand characteristics of a situation, Schulz von Thun (1998, 2004b, pp. 36-40) distinguishes between five situational components, namely, (a) situational roles and of the participants, (b) the network of projected goals and objectives, (c) the antecedents of a situation, (d) the subject matter or "thematic structure" and its (dis-)contents, and (e) the embeddedness of the protagonists in different systems.

Similarly, Euler (2004) examines (a) tasks and roles of the participants, (b) core characteristics of the situation, (c) procedures and structural courses of action, and (d) critical events. Interestingly, to illustrate the procedures and structural demand characteristics of socio-communicative situation, he explicitly chooses as an example the process stages found in many models of mediation. This further indicates that, that mediation, indeed, can be conceptualized as a specific case for which more general socio-communicative competence is required.

Contemporary personality process research also aims at understanding interpersonal situations beyond their surface characteristics by aiming at capturing their psychologically active ingredients across individuals (Mischel, 2004; Mischel & Shoda, 1995; Shoda, Mischel, & Wright, 2004). This is done by carrying out systematic research on the interplay between situational and intra-personal characteristics as will be further outlined below.

Intrapersonal Dynamic Processes.

Both the appraisal of the situation and the generation of behavioural responses are thought to be dependent on a complex process of intrapersonal processes and evaluations as elucidated in contemporary appraisal theories (Lazarus, Kanner, & Folkman, 1980; Frijda, 1993; Scherer, 1998) modelled in person systems theories (Cervone, 2005). One prominent meta-theoretical example is Mishel and Shoda's cognitive-affective processing system (CAPS) framework (Mischel, 2004; Mischel & Shoda, 1995; Shoda et al., 2004). Based on cognitive research outcome, the CAPS model suggests that features of situations are not just perceived and appraised both cognitively and affectively, but that they are also cognitively generated and that the intra-psychic situation (e.g. mood states and feelings) mediates the experience (Mischel & Shoda, 1995). According to the CAPS model, the personality system "contains mental representations consisting of diverse cognitive-affective units "…which include the individual's construal and representations of the self, other people, situations, enduring goals, expectations,-beliefs, and feeling states, as well as memories of people and past events" (Mischel, 2004, p. 11; see Table 2 below). This may also include the individual's personal values and principles, his or her *Weltanschauung* and general idea of Man (*Menschenbild*).

Table 2

Types of Cognitive-Affective Units

Encodings	i.e. categories (constructs) for the self, people, events and situations (externally and internally represented)
Expectancies and beliefs	about the social world about outcomes for behaviour in particular situations and about self-efficacy, about the self
Affects	i.e. feelings, emotions and affective responses (including physiological reactions)
Goals and Values	Desirable outcomes and affective states; aversive outcomes and affective states, goals, values and life projects
Competencies and Self-regulatory Plans	Potential behaviours and scripts that one can do, and plans and strategies for organizing action and for affecting outcomes and one's own behaviour and internal states.

Note. Cognitive-Affective Units are intrapersonal mental representations likely to be activated by internal and external situational appraisal (adapted from Mischel and Shoda, 1995, p. 253, Table 1).

One important regulatory cognitive-affective element found in Table 2 refers to the communicant's "judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986, p. 391), also termed perceived self-efficacy. Self-efficacy beliefs have been shown to affect cognitive, motivational, affective and selection processes which are highly relevant in situated social communication; among them are personal goal setting, task orientation, outcome expectancies, anxiety arousal, and career choice. Due to the centrality of perceived self-efficacy for communicative action, this concept will be referred to repeatedly in the following.

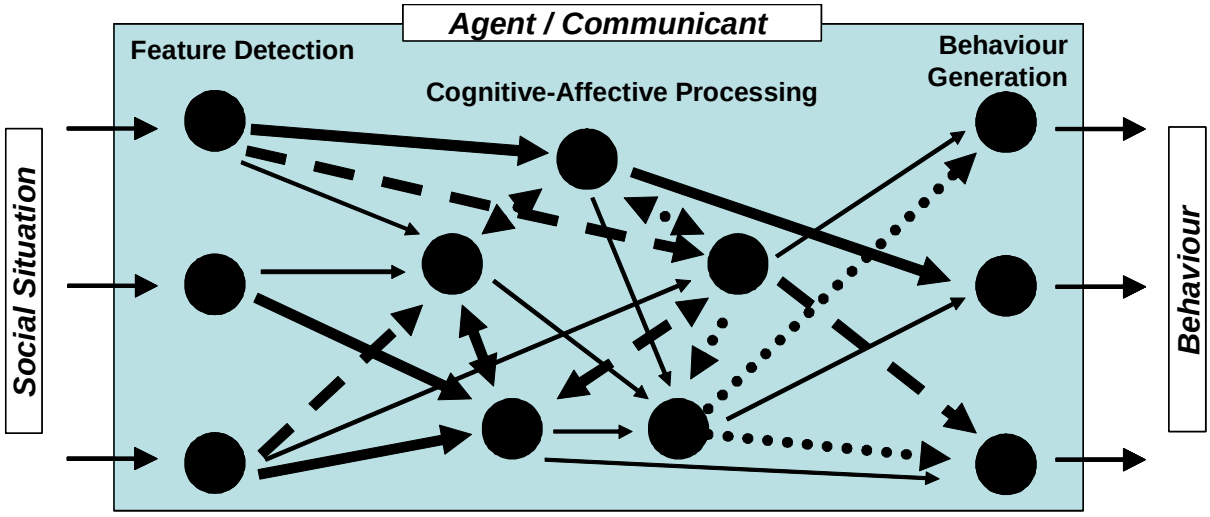


Figure 2. An illustrative CAPS personality system with a network of Cognitive-Affective Processing Units being activated by situational feature detection and associated processing units (adapted from Shoda et al., 2002, p. 318).

As Figure 2 indicates, these mental representations are seen as interconnected; for example, "thinking about a person can activate the memory of the thoughts and feelings associated with a particular event in the past, which in turn may lead to other memories and thoughts that may make us smile or cry …" (Shoda, LeeTiernan, & Mischel, 2002, p. 317).

Moreover, pairs of representations or "processing units" are thought to be characterized by a distinct and relatively stable strength of association between them.

Thus, in summary, an effective behavioural response for socio-communicative handling requires not only "situational feature detection" and behavioural capabilities or performance skills. As Schulz von Thun (1998, 2004a, 2004b; Schulz von Thun & Stegemann, 2004) has repeatedly pointed out, in order for a response to be effective, dual coherence (*doppelte Übereinstimmung*) is required. This means that any response should be in line with the nature of the external situation but, concurrently, with the nature of the internal state of affairs, that is: with the person of the communicant. It is seen as crucial that utterances and communicative behaviour is of authentic nature and in accordance with the communicant's individual and idiosyncratic identity. For example, the situational handling may essentially remain unrelated to and un-borne by the personality of the respective communicant when pre-scripted speech-balloons are practiced or when a communicant reproduces recommended or proclaimed behavioural responses: "The notion that, for every situation, a uniform ideal behaviour existed which every one should strive for, directly leads to a behavioural uniform, a 'communicative Sunday suit'…" (Schulz von Thun, 2004, p. 21). Instead, behavioural responses have to be adapted and carefully tailored to the "Inner Team" line-up which metaphorically describes both the internal variety and plurality of inner voices, their ambiguities, ambivalences, and their management. (The "Inner Team" will be discussed in section 2.2 in more detail as an ideographic strategy for tapping socio-communicative competence.)

Handling and Behavioural Response.

The generation of an idiosyncratic personal response to the situation is thus thought to be shaped and/or mediated by the individual's KSAOs. It is also assumed that the socio-communicative response generation is guided by a large number of behavioural scripts (Schank & Abelson, 1977) and cultural norms and references on various levels (general, societal, national, professional, group, etc.) of which an individual may be more or less aware. Through complex socialization and enculturation processes, various agents such as the family, school, peer groups, work, religion, and the mass media, are thought to shape, influence, and contribute to an individual's understanding of their response. Within the context of this discussion, two aspects are of importance to the subject of this thesis.

First, in contrast to many other types of processual knowledge of skills (e.g. solving mathematical equations or playing chess), the vast majority of humans are knowledgeable

'experts' in this field as they (can) draw on a multitude of experiences to handle novel situations that come about daily.

And, second, in most socio-communicative situations, an individual's espoused norms, values, and other systems of reference may seem inconsistent - or even paradoxical. They can, at best, only delineate a fuzzy array of conventional or acceptable behavioural responses. This implies that, theoretically, within a certain "bandwidth", an infinite number of behavioural responses to any given situation could be regarded as acceptable - if not even as "effective".

Socio-communicative competence: A working definition.

Summarizing the above, socio-communicative competence is understood as

- an array of individual knowledge, skills, attitudes, and other characteristics (KSAOs) (1) that shape and/or mediate
- an individual's appraisal or situational judgement (2) as well as the network of internal mental representations of
- the demand characteristics of specific types of professional interpersonal situations in which subject-/issue-focused social communication takes place (3)
- an effective response to or handling of which requires behavioural capabilities and/or performance skills (4) executed in dual congruence with the demands of the external situation but, concurrently,
- the individual's idiosyncratic, dynamic personality as explicated in its cognitive-affective internal situation (5) .

This working definition of socio-communicative competence will be used as a frame of reference in the remainder of this thesis. Its five essential elements are depicted in Figure 3.

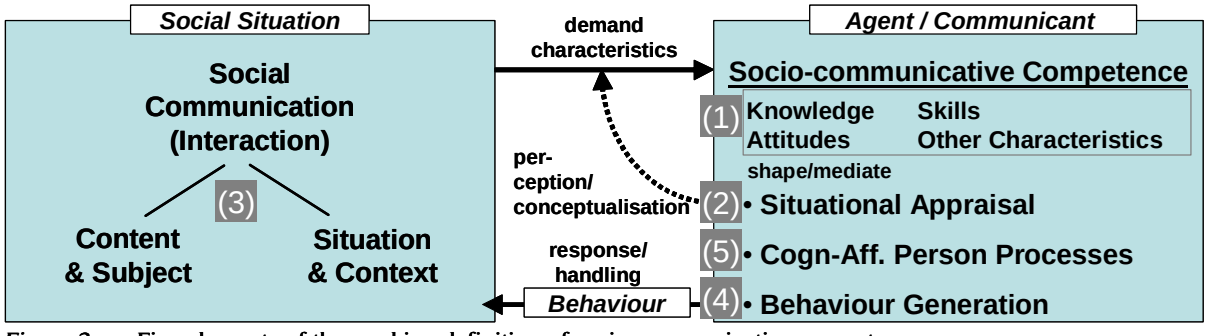


Figure 3. Five elements of the working definition of socio-communicative competence.

Socio-communicative competence in Mediation.

Yet, what are the specific demands of social situations in inter-personal mediation, and what are necessary prerequisites or KSAOs needed to effectively deal with these as a qualified neutral? Various research projects have attempted to provide more insight into this question.

In the late 1980s, the American *Society of Professionals in Dispute Resolution* (SPIDR) empanelled a Commission on Qualifications to examine the question of qualifications of mediators, arbitrators, and other dispute resolution professionals. In its first Report 1989, the Commission identified eight general skills seen as necessary for competent performance as a neutral third party across contexts and processes (see Table 3). In a subsequent survey among practitioners and programme administrators, a strong for these core skills was found. In addition, context-specific elements of competence were regarded as required to assist others in the resolution of disputes, such as sector or subject expertise (SPIDR Board of Directors' Commission on Qualifications, 1995).

Table 3

General Skills Necessary for Competent Performance as a Neutral

A. Ability to listen;
B. Ability to analyze problems, identify and separate the issues involved and frame these issues for resolution or decision making;
C. Ability to use clear, neutral language in speaking and (if written opinions are required) in writing;
D. Sensitivity to strongly felt values of the disputants, including gender, ethnic, and cultural differences;
E. Ability to deal with complex factual materials;
F. Presence and persistence, i.e., an overt commitment to honesty, dignified behaviour, respect for the parties and an ability to create and maintain control of a diverse group of disputants;
G. Ability to identify and to separate the neutral's personal values from issues under consideration; and
H. Ability to understand power imbalances.

Note. Based on SPIDR Board of Directors' Commission on Qualifications (1989), p. 17.

Subsequently, the mediation skill effectiveness framework pioneered by Christopher Honeyman and his colleagues at the Wisconsin Employment Relations Commission (Honeyman, 1988, 1990) and others (Honoroff et al., 1990) was used as a starting point for an independent research endeavour aimed at providing tools for the performance-based selection of mediators (see Table 4 below for an overview). The so-called *Test Design Project* consisted of various researchers and practitioners across the conflict resolution community. The *Project* followed up on the SPIDR Board of Directors' Commission on Qualifications' 1989 report, and it participated in editing the *Interim Guidelines for Selecting Mediators* (National Institute for Dispute Resolution, 1993) which had a significant impact on the mediation community.

Table 4

Early Framework of Parameters of Skills and Socio-communicative Effectiveness in Mediation

Investigation/Generating Information	Effectiveness in identifying and seeking out relevant information pertinent to the case
Empathy/Impartiality	Conspicuous awareness and consideration of the needs of others
Inventiveness and problem-solving/ Generating options	Pursuit of collaborative solutions, and generation of ideas and proposals consistent with case facts and workable for opposing parties.
Persuasion and presentation skills/ Facilitating Agreements	Effectiveness of verbal expression, gesture, and "body language" (e.g., eye contact) in communicating with parties.
Managing the interaction/ Generating Improved Relationships	Effectiveness in managing the process, coping with tensions and conflict between clients and professional representatives.
Strategic Direction	Ability to develop, set, and refine an effective, informed strategy of the mediation process.
Substantive knowledge	Expertise in the issues and type of dispute.

Note. Adapted from Honeyman (1988, 1990), Honoroff et al., (1990), and Honeyman & Pou, (1996).

Following the publication of the *Interim Guidelines*, fears were voiced that standardisation would adversely affect the variety of mediation and training programs (Menkel-Meadow, 1993; Silbey, 1993) and that it was likely to create a single model of practice which would then "be enshrined in court rules or legislation, and in turn imposed on programs whether or not they had values consistent with the Interim Guidelines' implied or expressed criteria of quality. (See, for instance, McEwen, 1993 and Pirie, 1994)" (The Test Design Project, 1995, p. 8). Ensuing these and other discussions on mediator qualifications (e.g. Bush, 1993; Dingwall, 1993; Duinker & Wanlin, 1994; Kolb & Kolb, 1993; Landau, 1994; LeBaron Duryea, 1994; Matz, 1993; Picard, 1994; Salem, 1993), the Test Design Project replaced the Interim Guidelines with its final report, *Performance-Based Assessment* (The Test Design Project, 1995). In the report, the authors conclude that "for anyone contemplating the introduction of any kind of standard, complicating factors abound; nearly every criterion of a mediator's job which has been articulated has also been disputed" (The Test Design Project, 1995, p. 4). They nevertheless maintain that mediator training and selection programmes should therefore clarify which values and criteria apply in the specific context, with close adherence to the culture in which the programme operates. Abstracting from a "common core" of behaviours which many mediators engage, the authors put forward initial lists of situational demands (tasks) and KSAOs which are "intended as a starting point to encourage any given program to prepare a modified list that reflects its actual practices" (The Test Design Project, 1995, p. 16). The lists are summarized in Table 5 below. It is this very idea of clarifying a mediation training programme's criteria that will be followed in this thesis.

Table 5

Lists of Situational Demand Characteristics (Tasks) and KSAOs in Mediation

Tasks / Situational Demands	KSAOs
A. Gathering Background Information 1. Read the case file to learn about the background and disputants. 2. Gather background information on a case from negotiators or other mediators (e.g. settlement patterns in similar cases). 3. Read legal or other technical materials to obtain background information. 4. Read and follow procedures, instructions, schedules and deadlines. ***B. Facilitating Communication*** 5. Meet disputants and make introductions. 6. Explain the mediation process to disputants. 7. Answer disputants' questions about mediation. 8. Listen to disputants describe problems and issues. 9. Ask neutral, open-ended questions to elicit information. 10. Summarize/paraphrase disputants' statements. 11. Establish atmosphere in which anger and tension are expressed constructively. 12. Focus the discussion on issues (i.e. not personalities or emotions). 13. Convey respect and neutrality to the parties. ***C. Communicating Information to Others*** 14. Refer disputants to specialists (e.g. alcoholism counsellors) or other services, or bring such specialists into the mediation process. 15. Refer disputants to sources of information about their legal rights and recourses. ***D. Analyzing Information*** 16. Help the parties define and clarify the issues in a case. 17. Help the parties distinguish between important issues and those of lesser importance. 18. Help the parties detect and address hidden issues. 19. Analyze the interpersonal dynamics of a dispute. ***E. Facilitating Agreement*** 20. Assist the parties to develop options. 21. Assist the parties to evaluate alternative solutions. 22. Assess parties' readiness to resolve issues. 23. Emphasize areas of agreement. 24. Clarify and frame specific agreement points. 25. Clearly convey to parties, and help parties understand, limitations to possible agreement. 26. Level with the parties about the consequences of non-agreement. ***F. Managing Cases*** 27. Estimate the scope, intensity and contentiousness of a case. 28. Ask questions to determine whether mediation service is justified or appropriate. 29. Ask questions to determine appropriate departures from usual practice for a given situation. 30. Terminate or defer mediation where appropriate. ***G. Documenting Information*** 31. Draft agreements between disputants.	1. **Reasoning**: To reason logically and analytically, effectively distinguishing issues and questioning assumptions. 2. **Analyzing**: To assimilate large quantities of varied information into logical ideas or concepts. 3. **Problem Solving**: To generate, assess and prioritize alternative solutions to a problem, or help the parties do so. 4. **Reading Comprehension**: To read and comprehend written materials. 5. **Writing**: To write clearly and concisely, using neutral language. 6. **Oral communication**: To speak with clarity, and to listen carefully and empathetically. 7. **Non-verbal communication**: To use voice inflection, gestures, and eye contact appropriately. 8. **Interviewing**: To obtain and process information from others, eliciting information, listening actively, and facilitating an exchange of information. 9. **Emotional stability/maturity**: To remain calm and level-headed in stressful and emotional situations. 10. **Sensitivity**: To recognize a variety of emotions and respond appropriately. 11. **Integrity**: To be responsible, ethical and honest. 12. **Recognizing Values**: To discern own and others' strongly held values. 13. **Impartiality**: To maintain an open mind about different points of view. 14. **Organizing**: To manage effectively activities, records and other materials. 15. **Following procedure**: To follow agreed-upon procedures. 16. **Commitment**: Interest in helping others to resolve conflict. 17. **Knowledge**: To have adequate legal or procedural knowledge of the subject matter in a specific situation (e.g. type of mediation program, state law). For some types of program little or no substantive knowledge is required prior to selection.

Note. Adapted from The Test Design Project (1995), pp. 16-18.

While the Test Design Project's publication remains an important point of reference for this and future research projects, other attempts to tackle the problem of defining KSAOs in the conflict resolution field have been undertaken. In the late 1990s, the *Mediator Skills Project* at the University of Georgia began its research collaboration with the major U. S. professional

organisations in the conflict resolution field to produce a job analysis relevant to third parties involved in interpersonal disputes. Job analysis has long been an important tool in industrial and organisational psychology for examining activities in a specific area of work, its context, and KSAOs needed for effective performance (cf. e.g. Sackett & Laczo, 2003).

The *Mediator Skills Project* asked divorce, family-child, and community mediators to complete "task diaries" on actual mediation cases that were content-analysed by the project, resulting in an extensive list of over 400 discrete tasks grouped in 14 functional categories, as well as catalogue of 18 knowledge and 13 skill categories (Herrman et al., 2001, 2002). Afterwards, both the task list as well as the knowledge and skill areas were submitted to large groups of practitioners for assessments of their relevance to practice. The mediators were asked to rate the importance and frequency of each task and, subsequently, the importance of a knowledge or skill for a given functional category. The results (as detailed in Table 6) clearly indicate that the most important tasks in mediation, indeed, are of socio-communicative nature.

Table 6

Empirically grounded Knowledge and Skill Areas of the Mediator Skills Project

Knowledge Areas	Skill Areas
Communication	Communication
Interpersonal Dynamics	Relationship Management and Encouragement
Ethical Issues	Ethical Issues
Mediation Process	Mediation Process Management
Information Gathering	Information Gathering
Personal Skills and Limitations	Mediator Error Correction
Solution/Agreement Formation	Dealing with Information
Conflict	
Problem Solving Techniques	Problem Solving
Power and Control	Power & Control
Inform/Educate/Disseminate/Teach	Education & Dissemination of Knowledge
Cultural Issues	Cultural Diversity and Competency Skills
Resources outside of mediation	
Administrative Practices/Procedures	Administrative
Alternatives to Mediation	
Theories of Social Change	
Mediation Models	
How to interact with involved people other than primary participants	

Note. Adapted from Herrman et al., 2001, 2002), ordered according to their relative importance (descending).

For example, "communication skills" refer to adjusting appropriate styles of communication

> "… to meet the needs of mediation participants; encouraging while structuring the flow of communication (e.g. actively listening, … asking questions …, … reframing, … using humor, metaphor, and narrative …); creating a safe environment for the expression of strong emotions; and using neutral language" (Herrman et al., 2001, p. 149).

Similarly, other important skill areas such as "relationship management and encouragement", "ethical issues", or "mediation process management" can be thought of as descriptions of socio-communicative handling competencies.

The knowledge areas listed above and their relative importance were used to develop a written test of 100 questions to assess mediator knowledge to be used, for example, in credentialing examinations. However, the authors conclude that their product does not test for skilfulness and that "a more complete package includes an application process and a test of skills to go along with a test of knowledge … . Creation of a full package is possible, but not a reality at this point in time" (Herrman et al., 2002, p. 45).

Socio-communicative Competence and Mediation.

In summary, the above discussion of the nature of socio-communicative competence yields four important aspects for the subject of this thesis.

First, it may be true that successful mediation requires many more KSAOs than those necessary to handle social communication. For example, knowledge of administrative procedures or computer skills are needed to adhere to standards of mediation documentation for court programs. However, albeit not sufficient, research indicates that socio-communicative competence may be not just a necessary, but even a principal ingredient in mediation.

Second, socio-communicative competence, in theory and practice, can also be seen as necessary in many other professional areas (such as practising law, psychotherapy or counselling, business consulting, or management). As such, socio-communicative competence may suitably be described as a meta-disciplinary or key qualification. Inversely, mediation can be regarded as an example for an array of professional situations a competent handling of which requires more general socio-communicative competence. However, as mentioned above, research indicates that (a) the central aspect of socio-communicative competence is interpersonal conflict management, and (b) the process stages in mediation are a prototypical example for situational features also found in other professional fields.

Third, it has been argued that effective mediation training must therefore include opportunities for the development of socio-communicative competence. However, the above also lends support to the notion that it may not be inappropriate to conceptualize mediation training not just as a special form but as an epitome (sic!) of a more general form of training for socio-communicative competence.

And fourth, research suggests that mediation training programmes – for purposes of credentialing and evaluation – need to identify the range of situative and cultural demands and the specific array of knowledge, skills, attitudes, and other characteristics intended to develop and/or assess in participants.

Consequently, as the subject matter of this thesis is the evaluation of the development socio-communicative competence in mediation training, two more general questions arise, namely:

- How can socio-communicative competence be assessed?
- How can socio-communicative competence development be promoted?

These questions will be explored in the following.

DIGEST 2.1

Socio-communicative competence is understood as an array of individual knowledge, skills, attitudes, and other characteristics (KSAOs) that shape the accuracy of a communicant's judgement or internal representation of specific types of professional interpersonal situations in which subject-focused social communication takes place an effective response to or handling of which requires performance skills executed in line with the individual's idiosyncratic and dynamic cognitive-affective state. Socio-communicative competence is a principal ingredient in mediation, and mediation training is viewed as one epitome of a more general form of promoting socio-communicative competence.

2.2 Approaches to the Assessment of Socio-Communicative Competence

Everyday Life.

In everyday life, as Schulz von Thun (1981, 2004b) notes, the course of further socio-communicative action generally depends on mutual reception, not utterance, and, generally, communicants are free to choose which aspect of the reciprocal utterances and communicative actions and they wish to attend to. The adequacy of a concrete communicative action is therefore regularly determined by evaluating its impact on the environment and the reaction of those involved and/or affected by the situation.

However, the idiosyncrasies of the socio-communicative discourse, the infinite opportunities for inter-punction, and the fuzzy bandwidth of possibly effective handling make systematic research difficult. To deal with this complexity, two distinct conceptual approaches have been taken (cf. Drasgow, 2003, pp. 124-127) which may also reflect the split between nomothetic vs. ideographic approaches to psychological research and assessment (Caprara, 1996; Cervone, 2005; Cervone, Shadel, & Jencius, 2001). These will be briefly sketched below.

Social and Emotional Intelligence Testing.

In one line of research, it was attempted to generalise or aggregate across situations and contexts by developing instruments explicitly intended as measures of social or emotional intelligence, or in the terms of Gardner (1983; 1993), *interpersonal* and *intrapersonal intelligence*. Examples are the *George Washington Social Intelligence Test* (Moss, Hunt, Omwake, & Ronning, 1927), Riggio's (1986) *Social Skills Inventory* (SSI), the *Social Competence Questionnaire* (Schneider, Ackerman, & Kanfer, 1996), or the *MSCEIT* (J. D. Mayer, Salovey, & Caruso, 2002, 2004). Tests of social and emotional intelligence have been found to correlate with personality tests (e. g. Davies, Stankov, & Roberts, 1998; Riggio, 1986; Schneider, Ackerman, & Kanfer, 1996). However, as Cervone et al. (2001) point out, this "top-down" approach to assessment aims at tapping global, phenotypic structural information based on differences between members of the groups studied. Yet, "pooling data from individuals may yield a simple between-person structure that does not capture the qualities of any of the individual persons (Molenaar et al. 2002)" (Cervone, 2005, p. 426). Individual differences approaches that essentially de-contextualise information (Karkoschka, 1998) therefore have little in common with assessing the effectiveness of professional interaction in specific

professional situations as in the working definition of socio-communicative competence outlined above.

The other line of research places emphasis on the specifics of the situation and the context. As such, these studies relate to the idea of "alternative assessment", i.e. that educational assessments should focus on measuring "genuine", complex skills in "authentic" contexts or real-world situations (cf. Gredler, 1996), preferably collected across various points in time to reflect individual progress (e.g. MacIsaac & Jackson, 1994). In sum, four main approaches to assessment in line with the working definition of socio-communicative competence can be conceptualised as attending to the elements shaped by an individual's KSAOs (see Figure 4 below):

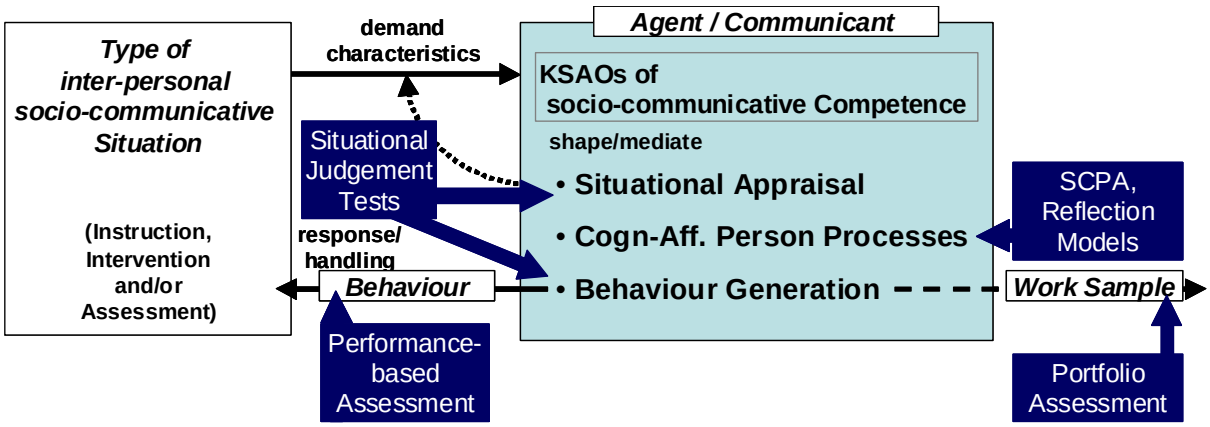


Figure 4. **Four approaches to the assessment of socio-communicative competence.**

- ***Situational Judgement Tests*** can be understood to capture a person's understanding of the situation and its (demand) characteristics as well as the suitability of generated behavioural courses;
- ***Performance-Based Assessment*** can be conceptualised to encompass an evaluation of a communicant's interactional behaviour and handling of the situational demands;
- methods used in ***Social-Cognitive Personality Assessment*** (SCPA) and contemporary communication models used for reflection such as the "***Inner Team***" metaphor (Schulz von Thun, 1998, 2004b; Schulz von Thun & Stegemann, 2004) aim at tapping and modelling an individual's cognitive-affective configuration and processing; and
- ***Portfolio Assessment*** aim at assessing the comparative qualities and unique characteristics of a number of work samples (such as learning diaries, essay-style conceptualisations of 'cases', situational demands, processes and/or behavioural

strategies, etc.) generally thought to be related to the development and level of socio-communicative competence.

In the following, these four approaches will be outlined further.

Situational Judgement Testing.

So-called Situational Judgement Tests (SJTs) have been used primarily to predict job performance. Here, an individual's understanding of the demands of particular situations is evaluated. The concept thus accentuates this very aspect or element of the working definition of socio-communicative competence exemplified above.

In SJTs, examinees are usually confronted with written or video-based depictions of situations supposedly representative of a specific target domain (e.g. a job). The examinees are then asked to respond to the situation; in most SJTs, by evaluating a list of possible responses to the situation.

As McDaniel and Nguyen (2001) observed, "there is substantial variation in how respondents are instructed to evaluate the potential responses to a situation" (p. 105). In some SJTs, the examinees are asked to identify the response they would "most" and/or "least likely perform" - which elicits self-reports and gives some indication that actual behavioural tendencies are the target construct. Depending on the test situation, this so-called "Behavioural Tendency" (Nguyen, Biderman, & McDaniel, 2005, p. 250) response format is suggested to be prone to faking and social desirability bias, however, especially in selection contexts (McDaniel & Nguyen, 2001; Nguyen et al., 2005; Peeters & Lievens, 2005). In terms of their psychometric characteristics, on the other hand, instructions to identify what one 'would do' displays more favourable characteristics than instructions in the those SJTs that ask respondents to identify the "best" and/or "worst response" to the situation presented, or to evaluate and rate all potential responses accordingly (Ployhart & Ehrhart, 2003).

By contrast, instructing examinees to identify what one 'should do' are "more clearly tapping knowledge of how to respond" (McDaniel & Nguyen, 2001, p. 105). As such, this response type is has been termed "Knowledge format" (Nguyen et al., 2005, p. 251) as it draws on the examinee's judgment rather than performance skill abilities, and probably these SJTs are more related to cognitive ability and job knowledge than to behavioural tendency

(McDaniel, Hartman, & Grubb, 2003; McDaniel, Hartman, Whetzel, & Grubb (III), 2007). This format is also considered more resistant to fake performance (Nguyen et al., 2005).

Traditionally, the test items are scored according to a key developed by the item creators (target scoring) or by use of subject matter experts (expert scoring) who had been asked to determine the "correctness" or "effectiveness" of the various response options. These scoring procedures may seem an appropriate procedure for instructional evaluation; for example, when the effectiveness of skill development (as projected by the authors of the educational programme) is being evaluated, student examinees could be asked in the aforementioned "knowledge format" to evaluate the potential responses in terms of what one 'should do' according to the authors of the educational programme. Waugh & Russell (2005) compiled 11 scoring algorithms typically used in these studies such as a simple count of correctly identified 'best' options ("pick-best") or correctly identified 'worst' options ("pick-worst"). More complex algorithms may include salient errors, for example, when the 'worst' alternative is incorrectly identified as the 'best' option ("pick-worst-reversal"), and/or vice-versa.

However, these traditional approaches to SJT scoring have recently been criticized on various grounds. Krokos, Meade, Cantwell, Pond, & Wilson (2004) argue that the complexity of social situations typically conveyed by situational judgement test items should preclude the idea that clear "right" or "wrong" answers could exist. Unsurprisingly, Lievens (2000) pointed out that consensus between subject matter experts in regard to the effectiveness of the various response options is often difficult to achieve. The obvious strategy in traditional test construction then is to exclude disputed items, which in turn, could lead to more transparent response options and, thus, diminished prospects to discriminate between high and low performers (Krokos et al., 2004). Moreover, expert judgments could be inferior to empirical scoring procedures in terms of their predictive validity, the provision of relative item weights, and the consideration of the interrelationships between items (Lievens, 2000). By contrast, empirical procedures could give more weight to response options that best differentiate between high and low performing examinees. Then, the "de-facto 'subject matter experts' are the high performing respondents as measured by the criterion of interest" (Krokos et al., 2004, p. 7). Initial findings suggest that empirical keying produces similar, if not better results in terms of reliability and predictive validity (Lievens, 2000; Waugh & Russell, 2005; cf. also MacCann, Roberts, Matthews, & Zeidner, 2004, for a related discussion).

SJTs generally display an acceptable criterion validity (ρ=.34 in McDaniel et al.'s meta-analysis, 2001), oftentimes high face validities (Motowidlo et al., 1990), and they tend to produce favourable applicant reactions (Clevenger, Pereira, Wiechmann, Schmitt, & Harvey, 2001). However, as Weekley and Ployhart (2005) note, at present, "there remains much uncertainty as to why they actually work and what they actually measure" (p. 82). Drawing on the results of their large-scale study they suggested that SJTs may be generally more reflective of cognitive ability and general work experience than of personality, training experience, or job tenure.

By drawing on the works of Clevenger (Clevenger et al., 2001) and Motowidlo and colleagues (Borman & Motowidlo, 1997; Motowildo, Borman, & Schmit, 1997), Bess (2001), suggested that SJT items may further be differentiated. Some test items may predominantly tap the knowledge of facts, principles as well as processing and decision-making procedures that are related to the functional or technical core of the target domain. This has also been called "task knowledge" by Motowildo et al. (1997, p. 80) and pertains to the "can do" or technical proficiency aspects needed to carry out the task (Chan & Schmitt, 2002). On the other hand, other test items may chiefly tap "contextual knowledge" (Motowildo et al., 1997, p. 80), which essentially relates to the "will-do" or knowledge aspects of performance beyond the core technical proficiency, i.e. motivational (e.g. volunteering, exercising self-discipline, goal-setting, etc.) and inter-personal performance (e.g. resolving conflict, working in a team, negotiating with others, etc.). However, when target domain of the task or "core proficiency" is essentially of an inter-personal or socio-communicative nature, as it is the case here, the task-context distinction may not make much sense.

An oral assessment form of situational judgment closely related to written SJTs is what is often referred to as structured situational or behavioural interviewing (e.g. Latham, Saari, Pursell, & Campion, 1980). However, due to the observable interaction between the interviewer and the interviewed, live or recorded situational interviews may additionally be analysed in terms of performance or impression management. In this respect they may also be thought of as a blend between content-related situational judgment testing and performance-based assessment which is discussed in the following.

Performance-Based Assessment.

In accordance with the working definition of socio-communicative competence, the individual's KSAO are also thought to shape the efficacy of handling or behavioural

performance. Thus, in order to assess situational performance rather than judgement, behavioural responses of the examinees must be focused on during assessment. Performance-based assessment is usually defined as a direct, systematic observation of actual examinee behaviours and the concurrent or subsequent rating of the observed performance (process) or the outcome (product) in accordance with criteria established previously. In assessment centres, for example, situations are created by realistic role-play simulation while the behaviours of examinees are observed and rated (cf. e.g. Jung, 2000). As such, performance-based assessment is a methodology based on systematic observation. In this very respect it differs from other means intended to give some indication as to the more general communicative performance, such as questionnaires or rating scales filled in, for example, by clients, partners, peers or, in self-assessments, the examinee him- or herself. (For an exemplary albeit generalizable overview of utilized assessment methods for communication skills in one professional area, namely, emergency medicine, see (Hobgood, Riviello, Jouriles, & Hamilton, 2002).

The Test Design Project (1995, see also Matz, 1993) put forward a methodology for the performance-based assessment of mediators by means of real-time case-based role-play simulations, followed by oral questions asked immediately after the role-play to assess behavioural motivations and cognitive conceptualisations. Performance-based assessment is also used in educational assessment and evaluation as an alternative to standardized tests (Wiggins, 1989). Performance-based assessment has also been discussed under the headings and conceptualizations of “real” or “alternative assessment” (e.g. Cumming & Maxwell, 1999), and “authentic assessment” (e.g. Meyer, 1992).

Initially conceptualized as paper-pencil procedures, both social intelligence and situational judgement approaches to socio-communicative competence assessment were considered standardized and economical (Bastians & Runde, 2002). By contrast, educational researchers have repeatedly criticised performance-based evaluation involving case-based simulations as expensive and uneconomical. The latter critique was underscored by Smit & van der Molen (1996) who undertook a comparative study of methods for the assessment of communication skills. They concluded that - while performance-based simulations may display a good face validity when based on different simulated cases for each examinee - this condition as well as the necessary and time-consuming preparatory, administrative and scoring procedures render its relative efficiency unsatisfactory. Another principal problem with

performance-based testing is its indefinite validity, both in terms of score generalizability (Linn, 1993; Parkes, 2001) and when bias-free consistency across examinees needs to be generated (Linn, Baker, & Dunbar, 1991). Moreover, performance assessments generally contend with poor inter-rater reliabilities (Bastians & Runde, 2002).

The Test Design Project (1995) also noted that "while role plays are the best method found to date for testing the skills at the core of a mediator's effectiveness, they are not the only, nor even necessarily the best solution in all situations" (p. 34). In their *Methodology* the *Project* explicitly refers to the practical problem that performance-based assessment by means of role-play simulation depends on a programme's "availability of in-house talent for drafting the selected exercise as well as for service as role-play actors and graders" (p. 7).

Another, more fundamental critique can be put forward when examining the implications of social-cognitive theory on assessment. As Cervone et al. (2001) noted, one must distinguish between overt behavioural or surface-level assessment and assessment of covert, underlying psychological dynamics: "The fact that people may exhibit similar surface-level profiles for entirely different reasons is well established, yet its implications for … assessment … have not been fully appreciated" (Cervone et al., 2001, p. 40). It is this fundamental difficulty that may have led the Test Design Project (1995) to include a post-role-play interview in which assessors can ask questions as to tap the underlying motivation of third party behaviours and the mediator examinee's "bent for self-awareness and self-criticism" (Matz, 1993, p. 328). This suggests that a sole performance-based approach to the assessment of socio-communicative competence may not do justice to examinees without tapping the underlying dynamics between individuals and the situation.

Tapping Internal Contingencies.

This very distinction between internal structure and overt behaviour is one prominent principle found in social-cognitive theories of personality dynamics and its assessment (e.g. Bandura, 1999, 2001; e.g. Caprara, 1996; Cervone, 2005; Cervone et al., 2001; Markus & Wurf, 1987; Mischel, 2004; Mischel & Shoda, 1995; Shoda et al., 2002; Zayas, Shoda, & Ayduk, 2002). Cervone et al. (2001) provide a detailed discussion of various other assessment principles derived from social-cognitive reasoning such as the call for an inclusion of measures to tap personal characteristics that may influence interpersonal action (e.g. beliefs, self appraisals and self-efficacy beliefs, personal goals, and self-regulatory skills). Consequently,

perceived self-efficacy is often used as an outcome parameter in evaluations of learning for socio-communicative skills as will be shown below.

Cervone et al. (2001) maintain that "assessing individual differences cannot substitute for assessing individuals" (p. 42), meaning that assessment methods employed should, in principle, be sensitive to the qualities of the unique individual – in terms of the cognitive content and the relationship between the personal constructs of an individual, but also the idiosyncratic situations that activate processing one important of which is perceived self-efficacy. Research based on these principles includes, but is not restricted to, measures that involve self-reports, diary writing, questionnaires, open-ended essays and free-response descriptions, categorization exercises, repertory grid, thinking-aloud methods, etc.

One prominent German communication model that seems to be of primary relevance to the issue discussed here is the "Inner Team" metaphor (Schulz von Thun, 1998, 2004b; Schulz von Thun & Stegemann, 2004). This metaphoric model -suggests that situational appraisal is generally accompanied by an assembly of contingent inner voices. In the Inner Team model, these voices are seen as representing inner messages, i.e. thoughts, feelings, values, states and traits, personality characteristics, etc. (In the social-cognitive conceptualisation of personality, these would probably be regarded as one or an assemblage of cognitive-affective units; see Table 2.) In similarity to work teams and social groups operating in external reality, inner teams are seen as being composed of a (often limited) number of lined-up voices or messengers. Their messages can be tapped by self-reflection and be visualised externally, for example, in drawings. Used as a method in training, supervision and coaching contexts, these drawings often entail, within a draft upper part of the protagonist's body, sketches of the various inner voices with their messages embodied in the imaginated and sketched appearance and posture as well as in a speech-bubbled key phrase. Depending on the elaborateness of the drawing, the voices' perceived relative internal strength, position, and the network of – conflicting, supportive, and neutral – relationships can also be expressed.

Portfolio Assessment.

In many fields as, for example, writing, visual arts, crafts or even music, the various "products" created or generated by an individual across time may be analysed. Work samples thus provide a basis for an assessment of skill levels (e.g. Wolf, 1989). In higher education contexts, portfolio assessment is often viewed as a means to evaluate an individual's skills for reflection and critical thinking (e.g. Jarvinen, 1995). Within this postsecondary educational

context, however, portfolios have mainly been discussed as a means to teacher development and reflective practice, and less for learning (Klenowski, Askew, & Carnell, 2006), for example, by use of reflective journals, learning diaries, and the results of written tasks and exercises (cf. Krainz, 2005), but portfolios may also comprise video artefacts of interpersonal encounters (Hobgood et al., 2002) and collections of oberservations and reflections in cross-cultural contexts (Jacobson, Sleicher, & Maureen, 1999). Theoretically, these may also be used for further assignments, such as, for example, the production of cultural assimilators or sensitizer items (Fiedler, Mitchell, & Triandis, 1971). The quality of these 'derivative' work samples should subsequently be rate- or gradable and may be taken as one indicator of socio-communicative competence of the assessed. (Possibly, the culture assimilator concept may also be transferred from the inter-cultural context to same-culture types of professional enculturation and role-bound social skill development.)

Unlike knowledge and practical skills, socio-communicative competence is essentially situationally and subject- or issue-bound, and, thus, requires much contextual information. For these reasons, assessment of ultimate behavioural performance, context-induced cognitive-emotive processes as well as attitudes and situational judgment are principally preferable to assessing "derivative" work samples which are chiefly created either without the assessor's knowledge of the situation or artificially, i.e. for assessment purposes only. As such, this approach to assessment is not truly situated, which is a definite strength of the other three approaches to the assessment of socio-communicative competence. In addition, the provision of evidence that the (rated) quality of unique, ideosncratic work samples is a valid indicator of socio-communicative competence (or even a valid predictor of competent behaviour) may also be more difficult.

Assessment Purposes and Applicability.

The working definition of socio-communicative competence may suggest that combining (at least the first three of) the aforementioned “situative” approaches could yield a most advantageous research strategy. However, the comparative relevance of the aforementioned approaches to socio-communicative competence assessment depend not just on their relationship to the working definition, but also on the very objective of assessment. For example, assessment can be used for purposes of selection, placement, certification, credentialing, and licensing; it may support learning, training and instruction, guidance and

counselling; it may assist in human research performance reviewing, and it may serve as a method of systematic research and evaluation.

The subject matter of this thesis pertains to two distinct target domains or domains of inference. One the one hand, as this study's setting is a specific mediation training programme (see below), socio-communicative competence development is deliberate and specifiable in terms of a "curricular domain" (cf. Millman & Greene, 1989, p. 336) within which summative evaluative inferences shall be made about the effectiveness of the programme. Yet, on the other hand, this study seeks to explore instructional effects beyond the instructional curriculum which relate to more theoretical perspectives on computer-supported communicative competence development. Traditionally, in functional descriptions on assessment, this purpose is – somewhat misleadingly – referred to as the "cognitive domain" (cf. Millman & Greene, 1989, p. 336) as most educational research stimulated by theory and practice is concerned with cognitive knowledge, processes, and skills.

The choice of assessment type therefore depends on its applicability to serve both functions. As data will have to be aggregated in both cases, these purposes generally require a comparability of assessment-generated information between individuals and across time, i.e. assessment instances. It follows from the discussion above that situational judgement tests generally seem best suited for this purpose while uniquely individualized approaches may yield a great variety of accounts of situative individual contingencies or cognitive-affective processing. As open-ended response formats generally make comparisons between individuals difficult, quantitative and comparative research usually relies on more classifiable self-reports. Performance-based assessments seem appropriate in case the problem of inter-rater reliability can be tackled and a good level of situative consistency across instances and examinees and can be generated without losing too much authenticity. As will be outlined in the following, computer technology-may be a means of sophisticating these three variants of assessment tasks relevant in the socio-communicative context.

Approaches to Technology-Supported or "Blended" Assessment.

Recently, multi-dimensional frameworks have been put forward to structure the discussion how technological innovations can support or replace traditional approaches to assessment (Bunderson, Inouye, & Olsen, 1989; Parshall, Davey, & Pashley, 2000; Zenisky & Sireci, 2002). While Bunderson, Inouye, and Olsen (1989) discussed computerized testing chiefly as an alternative to and improvement of traditional educational measurement, Parshall

et al. (2000) attempted to categorise test features that are altered by and/or "blended" with computer technology. Their framework (p. 130) covers five inter-related dimensions (item format, response action, media inclusion, level of interactivity, and scoring algorithm, along which modifications of "traditional" situational judgement, performance-based assessment, and internal contingency tasks) and will be outlined in the following.

(1) Traditional (paper-and-pencil) item formats consisted of either constructing a response (as in open-ended essay question) or selecting from a number of pre-constructed responses (as in multiple-choice questions). Item formats used in computerized or computer-assisted assessment, are much more diverse. For example, Zenisky and Sireci (2002) compiled a list of 21 novel item formats, including so-called drag-and-drop (examinees click and drag an object to the centre of an appropriate answer field), passage editing (examinees edit text passages by selecting rewrites from a list of alternatives in a drop-down menu), or simulation or role-play items (examinees react as a dynamic situation unfolds).

(2) Technology has also made possible novel interactions with the test taking media, (i.e. the computer). In high similarity to traditional paper-and-pencil exams, the examinees' *response actions* could include typing text-based or entering numerical answers by means of a keyboard; however, drawing on screen, clicking buttons or prompts, the utilization of light pens, joysticks, trackballs, microphones combined with speech recognition software, or pressure-feedback (haptic) devices represent novel ways of technology-assisted data collection.

(3) *Media inclusion* refers to the use of graphics, video, or sound within an item stem. While traditional paper-and-pencil tests are essentially restricted to low-fidelity formats, technology-assisted assessment may provide for high-fidelity testing through the use of graphics (i.e. in medical diagnostics; cf. Ackerman, Evans, Park, Tamassia, & Turner, 1999), audio (Parshall & Balizet, 2001), and/or video sequences (Chan & Schmitt, 1997; Drasgow, Olson-Buchanan, & Moberg, 1999; Frederiksen, Sipusic, Sherin, & Wolfe, 1998; Funke & Schuler, 1998; Havighurst, Fields, & Fields, 2003; Hulsman, Mollema, Hoos, de Haes, & Donnison Speijer, 2004; Humphris & Kaney, 2000; Lievens & Sackett, 2006; Runde, Bastians, Kluge, & Wübbelmann, 1999; Schoech, 2001; Schuler, Diemand, & Moser, 1993; Smit & van der Molen, 1996; Weekley & Jones, 1997). Interestingly, recent research suggests that video-based versions of an interpersonally-

oriented SJTs likely have lower correlations with general cognitive ability than written versions. They also may have higher predictive and incremental validity for predicting interpersonally-oriented criteria than written versions (Lievens & Sackett, 2006).

(4) Technology-assisted assessment also offers higher *levels of interactivity* between examinees and the test content (so-called content interaction, cf. Northrup, 2002) and, thus, higher levels of individualized test-taking[4]. When examinees take multimedia tests at their own pace, for example, this may in turn result in higher levels of administrative efficiency and measurement accuracy (Vispoel, 1999). Branching in computerized assessment scenarios (i.e. in counselling skills assessment; Sharf & Lucas, 1993) integrates adaptive testing and simulative experience. In accordance with Haack (2002), progressively higher levels of content interaction are achieved by implementing the following characteristics: (a) access to specific informational content, (b) links to respective additional information, (c) highlighting (i.e. colouring) pieces of information and activating relevant additional information, (d) the provision of intelligent or adaptive feedback on the spot. Another new facet of content interactivity lies in the inclusion of technology–based or –supported or –controlled access to reference materials (e.g. electronic books, dictionaries, or other media) and applications (e.g. electronic calculators, language translation devices, or other software) examinees may use as they take the test (Zenisky & Sireci, 2002).

(5) Finally, technology-supported solutions enhance the ways in which responses can be scored and used in testing (Bergstrom & Lunz, 1999; Bunderson et al., 1989; Drasgow, 2002). In contrast to traditional paper-and-pencil testing, computers allow for the use of complex *scoring algorithms* on-the-spot, and thus, renders possible complex computerized-adaptive testing, the scoring of open-ended responses (cf. Zenisky & Sireci, 2002, for a summary of automated essay scoring procedures), and the provision of individualized feedback (see also Azevedo & Bernard, 1995).

In future, technology- or computer-supported assessment is likely to play a major role in the assessment of socio-communicative competence. Early approaches to video-based assessment include technology-based situational judgement tests employed in the hotelling (C. Jones & DeCotiis, 1986), transport (Smiderle, Perry, & Cronshaw, 1994), insurance (Dalessio,

[4] While interactivity *between* test-takers or *social interaction* may not play a prominent role in indidvidual computer-supported assessment, technology also provides solutions for the 'blended assessment' of interactions.

1994), and retail (Weekley & Jones, 1997) industries and in child protection services (Schoech, 2001). Less specific in terms of industry-or job-relatedness were tests intended to select customer service personnel (McHenry & Schmitt, 1994).

The aforementioned examples pertain to the use of technology in the assessment of more or less job-specific competencies. In the following, approaches to technology-supported assessment of communication skills and, more specifically, conflict resolution skills will be detailed.

In what Drasgow et al. (1999) characterized as "one of the most carefully developed video assessments" (p. 178), Desmarais et al. (1994) used a SJT approach in which video-based vignettes depict vocational interpersonal issues relating to "poor training, demanding workloads, interpersonal conflict, sloppy work habits, and flawed work". A similar approach based on critical incidents in personnel management was taken by Runde et al. (1999) who produced video-, audio- and text based items for a traditional situational judgement test of (unspecified) social competence. Interestingly, examinees here were asked to opt for only one "would-do" response alternative. Results indicate moderate concurrent correlations to the ratings of test participants in performance-based assessment-centres.

A computerized adaptive test of conflict resolution skills was developed by Olson-Buchanan, Drasgow, Moberg and their colleagues (Drasgow et al., 1999; Olson-Buchanan et al., 1998), utilizing videos in a SJT format and branches dependent on the examinees' responses. Their findings further suggest that psychometric analyses here could be more similar to the analysis of biographical data inventories than to traditional academic skills tests (cf. also Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004).

More recently, Bakx, Sijtsma, Van der Sanden, and Taconis (2002) developed a multimedia assessment instrument for social-communicative competence in social work. The instrument consisted of a series of tests for basic communication skills, 'bad news' dialogues, advising clients, and counselling and entailed video fragments with subsequent questions. Their evaluative report suggests that training (or increased expertise) may result in better performance as fourth year students scored significantly better than second-year students which, in turn, obtained higher scores than first-years.

For evaluation purposes in medical education, Humpris and Kaney (2000) asked medical undergraduates to complete a test intended to measure their capability to directly

identify and discuss communication skills displayed by clinicians in video-based interactions with their patients. This approach likely assesses predominantly (course-based) knowledge, however, and thus, seems somewhat inept to serve as a model for this study. Similarly, Hulsman and colleagues (2004) created a computerized video-based version of the so-called "objective structured clinical examination", a performance-based testing device used in the UK and at the Academic Medical Centre in Amsterdam, The Netherlands. Again, by subscribing to a situational judgement test format with open-ended questions, the authors intended to assess communication skills such as "exploring patient ideas, concerns and expectations" during medical history taking, "breaking bad news", and "neutral presentation of options" in treatment decision making. Their evaluations suggest that - despite the fact that the video assessment was generally favourably received by students – the time required for manual rating of open-ended essay questions is likely to render this response format impractical for large-scale evaluation.

In summary, high-fidelity, technology-supported, video-based situational judgement testing likely is an often opted-for, feasible approach in the study of socio-communicative competence which used for both assessment and evaluation purposes. Recent research suggests that video-based situational judgement testing may profit from higher predictive and incremental validity for predicting interpersonally-oriented criteria than written tests, while keying and scoring procedures are still a matter of intense debate.

DIGEST 2.2

In accordance with the working definition of socio-communicative competence, its assessment focuses on the evaluation of the idiosyncratic ways an individual perceives situational demands, the tapping of idiosyncratic internal responses, and an evaluation of the respective behavioural outcome. While general tests of social intelligence lack situational specifity, both Situational Judgement Testing and Performance-Based Assessment seem generally appropriate methodologies for the purpose of this study, especially when they have the potential to capture essential processing characteristics of the unique individual. Recent developments in technology-supported or "blended" assessment render high-fidelity, video-based situational judgement testing a feasible method for the study of socio-communicative competence.

2.3 Perspectives on the Development of Socio-Communicative Competence

In order to evaluate the development socio-communicative competence in mediation training, one needs not just an assessment methodology, but also a developmental model that guides practice, research, and evaluation. The subject under study here pertains to two main questions, namely, (1) how socio-communicative competence is acquired and what methods facilitate development processes and (2) if the developmental processes can be supported by technological means.

Learning theories account for variables which influence learning processes and provide explanations of the ways in which these influences occur. They refer to epistemology, i.e. the historical and philosophical nature of knowledge, and make fundamental assumptions about what is crucial for understanding, designing, and evaluating learning and instruction, and designate the conditions and factors mediating learning and cognition. Underlying instructional and learning theories also play a significant role in the evaluation of instructional and learning systems since modern evaluation theories call for the introduction of existing knowledge and theory into evaluation design (e.g. Rossi, 1999; Shadish, Cook, & Leviton, 1995).

Therefore, it seems a worthwhile effort to review major perspectives on learning and instruction in regard to three aspects, namely, (a) how the development of socio-communicative competence can be conceptualized, (b) what perspective-based shapes e-learning-technology may assume in learning and instruction and (c) how assessment and evaluation as viewed from the respective perspective.

Major Perspectives on Learning and Instruction.

Greeno, Collins, and Resnick (1996) grouped philosophical traditions and theoretical and research contributions in this field into three broad categories, referred to as behaviourist-empiricist, cognitive-rationalist, and situative-socio-historic. In the following brief review, their general distinction will serve as a model to start from in trying to determine and place into perspective the theoretical background of this study. However, for the purposes of this study, it seems warranted to depart from their general classification in one noteworthy aspect. While Greeno et al. (1996) subsumed constructivism under the heading of cognitivist-rationalist theories, it might very well be regarded as the leading metaphor of human learning today (cf.

Mayer, 1996), and it seems of particular importance for the e-learning domain. Indeed, as Driscoll (1994) points out,

> "it is probably no accident that constructivism is gaining popularity and momentum at the same time as interactive, user friendly computer technologies are becoming widely available. The computer offers effective means for implementing constructivist strategies that would be difficult to accomplish in other media." (p. 376)

Thus, four perspectives will be reviewed in the following.

Behaviourist-Empiricist Conceptualisation.

In the traditional behaviourist perspective, learning can be conceptualised as the formation, strengthening, or the adjustment of associations (Carlile & Jordan, 2005; Greeno et al., 1996; Schulmeister, 2002a; Tergan, 2004) between external environmental conditions -which offer antecedents (stimuli) and consequences - and the (behavioural) responses of the learner. In line with on empiricist philosophies, knowledge is viewed as achievable and measurable in that the responses of an individual can be meet objective, external criteria to varying degrees.

Pedagogical approaches originally based on behaviourist concepts (such as Programmed Instruction) emphasise aspects of practice generally found to be effective. Issues of practical conceptualization include repetition in learning, the provision of adequate (strong and varied) stimuli, a careful planning and sequencing of learning events, specifying objectives of learning, grouping them into manageable steps and sequences, providing instructional hints or cues to guide learners towards the desired behaviour, and reinforcement principles (e.g. immediate feed backing, intermittent reinforcement).

One of the key paradigms of technology-based learning that is situated chiefly within the behaviourist-empiricist perspective is *Computer-Aided Instruction* (CAI) (Koschmann, 1996, 2001). CAI – either as support or replacement of traditional forms of instruction - is widely used for the routinized transfer or consolidation of well-structured pieces of knowledge. CAI is associated with rather simple, linear "drill and practice" schemes in which the learner's responses can be evaluated in terms of "right" or "wrong" (Sembill & Wolf, 2001).

CAI-based learning systems – as, indeed, behaviourist models in general – therefore seem of insufficient applicability to higher level learning (cf. Carlile & Jordan, 2005; Sembill & Wolf, 2001; Tergan, 2004). Indeed, if – as in the working definition of social competence -

knowledge, skills, attitudes and other characteristics are viewed as a mediating factor in perception, understanding, construction and/or conceptualisation as well as internal and external management of the complex demands of social communication, this well surpasses the distinct and hardly standardizable contingencies between the ever changing social situation and the behavioural responses of the learner.

Cognitivist-Rationalist Conceptualization.

Cognitivist-rationalist concepts can be viewed as historic responses to the limitations of exogenic (or externally -centred) behaviourist models. In contrast to the latter (which have primarily focused on observable environmental and behavioural phenomena), cognitivist-approaches maintain that environmental "cues" and instructional components alone cannot satisfactorily explain all the learning that results from an instructional situation (Ertmer & Newby, 1993).

Cognitivist approaches follow rationalist traditions in that they emphasize *endogenic* (or mind-centred) orientations to knowledge (cf. Gergen & Wortham, 2001): Foci of cognitivist investigation are endogenic human thought processes (such as memory, perception, or attention), cognitive abilities (such as reasoning, planning, or problem-solving), and cognitive and emotional states relevant to learning (such as beliefs, attitudes, or motivation). Various lines of research were promoted by theories of development (e.g. Inhelder & Piaget, 1958), Gestalt theory and problem-solving (e.g. Newell & Simon, 1972), and linguistics (e.g. Chomsky, 1959).

Within this perspective, learning essentially is understood as a change in conceptual and cognitive structures, i.e. e.g. mental representations, thinking strategies, conceptual understanding, meaning, or the conditional, declarative, and procedural understanding of phenomena (cf. e.g. Schunk, 2004). "Promoting mental processing" instead of "promoting overt performance" also created a shift from behaviourist procedures for manipulating the learning materials to cognitive procedures for directing student processing and interaction with the instructional system (Ertmer & Newby, 1993). Memory is given a prominent role in the learning process: Cognitivist pedagogies underscore the readiness and active involvement of learners, the importance of linking new knowledge to existing cognitive structures to make it meaningful and to increase the probabilities of long-term retention. Genuinely, cognitive pedagogies focus on changing the learner through encouraging him/her to utilize appropriate learning strategies (Ertmer & Newby, 1993).

Cognitivist-rationalist conceptualizations retain the traditional proposition that knowledge is essentially in the "possession" of individuals, and that it can – and should –be measured. Pedagogical principles originating from behaviourist theories – such as the importance of feedback, repeated practice, and environmental cueing – are also stressed by cognitivist-rationalist approaches to practice, albeit for different reasons (cf. Schunk, 2004). For instance, from the behaviourist perspective, feedback is seen as reinforcement and targets at modifying behaviour into the direction desired by the instructor; from a cognitivist perspective, feedback is viewed as guidance in that information is disclosed to the learner about the degree to which an instructional objective has been accomplished. Here, feed backing targets at supporting accurate mental processes (A. D. Thompson, Simonson, & Hargrave, 1992; cit. in Ertmer & Newby, 1993).

Drawing on artificial intelligence and information-processing research, cognitivist instructional designers sought new ways of employing technology for instruction - maybe best embodied in the Intelligent Tutorial Systems (ITS) paradigm (Koschmann, 1996, 2001). ITS are interactive, automated instructional systems oriented towards an adaptive dialogue with the learner: "Knowledge about a complex subject is offered and questions are asked. New information or already introduced units are provided on the basis of the learner's answer." (Sembill & Wolf, 2001, p. 74).

Excursus: Transmission Models.

Beginning in the 1950s, educational researchers from both the behaviourist and the cognitivist perspective provided material for what have become influential systems for curriculum planning and the delivery of instruction (e.g. Ausubel, 1968; Bloom, Englehart, Furst, Hill, & Krathwohl, 1956; Gagné, 1965; Gagné & Medsker, 1996). For instance, one prominent concept rooted in the ISD tradition of analysing the conditions of learning is Gagné's specification of learning outcomes and learning events (cf. Gagné, Briggs, & Wager, 1992). According to this model, learners acquire capabilities that are manifested in five types of learning outcomes, i.e. (1) intellectual skills (or procedural knowledge, employed e. g. in writing or reading), (2) cognitive strategies (or problem-solving, employed e. g. in combining or reprocessing information), (3) verbal information (or declarative knowledge), (4) motor skills, and (5) attitudes. In modern approaches to so-called Instructional Systems Design (ISD), the entire range of tailoring content and delivery in instructional systems is covered (Reinmann-Rothmeier & Mandl, 1997). ISD can be understood as an endeavour to optimise the

assimilation of subject matter (i.e. the exogenic, external, environmental or curricular conditions of learning) and to concurrently consider the accommodative needs of the learners (i.e. the endogenic, internal, cognitive, and individual conditions of learning). To achieve these objectives, ISD often includes an analysis of learners and their requirements, an analysis of the target group and the conditions of learning, the establishment of instructional goals, the selection of learning content, the planning of learning methods, didactics, and delivery formats, the development and application of learning materials, and evaluation and revision procedures (Issing, 2002). (A contemporary review of other instructional systems rooted in the behaviourist tradition such as PSI, Precision Teaching, or Direct Instruction and their relationship to instructional technology can be found in Burton, Moore, and Magliaro, 2004.)

Instructional transmission models implicitly endorse the "Mind-as-Container" (or, similarly, "Mind-as-Computer") metaphor (Lakoff and Johnson, 1980; cf. also Bereiter, 2002; Duffy & Cunningham, 1996). which effectively regards (a) knowledge as a set of specifiable objects (e.g. discrete facts, beliefs, ideas, or intentions) that are seen as accurately relating to external world and (b) an individual's cognitive skills in terms of doing specifiable things (e.g. classification, sequencing, retrieval, reasoning) with specifiable mental objects (Bereiter & Scardamalia, 1996). These cognitive objects or symbols are essentially viewed as abstract and independent of individual experience (Duffy & Cunningham, 1996). In accordance with the Mind-as-Container-metaphor, educational testing can then be understood as "inventorying" cognitive content (Bereiter & Scardamalia, 1996).

As models of "transmission", ISD systems have been portrayed as being based on the tacit assumptions that learning involves the accumulation of particular sets of (objective) facts and (external) skills, and that instruction involves the transmission of these from experts (teacher) to individual novice students (learner). In other words, the transmission goal of instruction is to map the structure of the world onto the individual learner (Jonassen, 1991; cit. in Ertmer and Newby, 1993).

Viewed from this angle, approaches to blending traditional face-to-face with technology-based instruction should then attempt to make this transmission more efficiently, either in terms of outcome or cost. For example, the aforementioned technology-based Intelligent Tutorial Systems (ITS) try to emulate skilled instruction by using ISD algorithms aiming at representing pedagogical and domain knowledge to enhance student understanding (Koschmann, 2001).

Instructional transmission models, in accordance with the Mind-as-Container-metaphor, regularly view assessments as both feasible and warranted (Cognition and Technology Group at Vanderbilt, 1996; Greeno et al., 1996). Critics argue that assessment often only entails recalling appropriate formulas, and making substitutions to get "correct" answers. Consequently, they argue, students conceive knowledge as collections of facts and tend to rely on learning strategies other than those required for "deep learning" (Gipps, 2002). This may be one reason why critics point to possible deficiencies of cognitivist approaches in various learning domains such as the acquisition of cultural techniques or the development of key qualifications (e.g. media and ICT skills, communication and social skills, learning and problem solving skills) (Reinmann-Rothmeier, 2003b, p. 37). Indeed, especially in the socio-communicative domain, focusing only on the transmission of surface-level knowledge may prove fruitless. For example, a simple lecture on communication models may neither effect changes in situational judgement and appraisal, nor in internal cognitive-affective responses nor in the behavioural handling of communicative situations in the addressees. However, as will be outlined further below, social-cognitive learning theories provide detailed explanations of active or facilitative "ingredients" thought to promote the development of situational judgement and appraisal, the inclusion of internal cognitive-affective responses and the behavioural handling skills.

Another critique of instructional transmission models pertains to the perception that these systems inherently focus on instruction rather than learning and thus are not learner-based. They have been considered to purport a 'nutritionist' model (Friere, 1985) which is "… essentially hierarchical, with the ultimate authority residing in the communities of knowledge-production" and with students, being 'fed' or consuming the "educational nutritients" at the lower end of the hierarchy (Gergen & Wortham, 2001, p. 125). However, it may be argued that the development of socio-communicative competence may not be dependent on formal educational settings as not so much knowledge as appraisal is at the heart of socio-communicative competence (cf. Lazarus & Smith, 1988). Theoretically, any social situation can be used for learning purposes, especially from an agentic perspective of personhood and personality (Bandura, 2001; Mischel, 2004). Moreover, experiential and skill-based instruction may – in contrast to the argument presented above – still be regarded as neglected fields in university-based higher education. And, finally, voluntary vocational training of motivated practitioners in many social professions (including mediation) is essentially

learner-based and entails opportunities socio-communicative competence development. However, the level of professional socio-communicative competence is neither dependent on nor guaranteed by training. It is precisely because of this implication that, for the field of mediation, the SPIDR Board of Directors' Commission on Qualifications' (1989) has called for performance-based testing of applicants and not reliance on educational certificates.

Constructivist Conceptualisations.

As cognitive theorists started to focus on learning processes rather than instruction, emphasizing that the learner and his/her ways of experiencing and creating meaning are at the heart of the generation of knowledge (rather than instructional methods or media), a new educational paradigm started to emerge (Cooper, 1993; Jonassen, 1991; Cognition and Technology Group at Vanderbilt, 1996): *Constructivism.* Constructivist thinking, with its roots traceable back to Piaget (e.g. Inhelder & Piaget, 1958), fundamentally departs from the transmission model outlined above in nearly every single aspect and is based on a different set of (equally) untestable axioms.

Firstly, in contrast to instructionist models, constructivists view humans as not capable of transcending their perceptional and linguistic capabilities. In essence, the realm of human cognition is seen as comprised of idiosyncratic propositions based on observational or experiential processes. These propositions are understood to be tied inextricably to the person observing: 'Everything said is said by someone' (cf. Maturana, 1982). Learning, then, is a process of actively *creating meaning* (from experience) rather than *acquiring* it (Ertmer & Newby, 1993). These axioms imply that the individual learner knows only what he or she has constructed and that it is essentially impossible to "know" what someone else has constructed (Ludewig, 2002).

Secondly, when a multiplicity of possible and idiosyncratic meanings can be constructed, the ideas of (external, ontological, or instructional) 'objectivity', 'truth' or 'certainty' implied in both empiricist and rationalist models become obsolete or irrelevant. In contrast to more traditional conceptions, knowledge is not seen as having the purpose of producing "representations of an independent reality, but instead has an adaptive function" (von Glaserfeld, 1996, p. 3). The concept of 'objectivity', therefore, is substituted by 'viability' (Duffy & Cunningham, 1996): Correspondences between different observations and descriptions may be pragmatically useful for purposes of communication. In other words, "any knowledge to be constructed has to be viable for its agent [i.e. the learner] under the particular

conditions of the case" (Karagiorgi & Symeou, 2005, p. 18). Knowledge, then, "a consensus of beliefs, a consensus open to continual negotiation (Rorty, 1991)" (Duffy and Cunningham, 1996, p. 10) and to personalisation of meaning and relevance.

Thirdly, learners, as people in general, are essentially understood as un-instructible (cf. Ludewig, 2002). Learning, on the other hand, is not seen as the *result* of development; it is seen *as* development (cf. Fosnot, 1996, p. 29). "Instruction" then, is a term that is fundamentally questioned by constructivists, and, as Duffy and Cunningham (1996) noted, understood as processes supporting the construction of knowledge and/or meaning. Constructivists presuppose that every learner has a unique, autonomous perspective and are interested in the learners' a explorations of his or her environment and their skills of reflexivity rather than of remembering (cf. Bednar, Cunningham, Duffy, & Perry, 1995). In accordance with the ideas of viability and knowledge creation, constructivist pedagogies subscribe to the idea that "disequilibrium facilitates learning", meaning that "challenging, open-ended investigations need to be offered, … allowing learners to explore and generate many possibilities … . Contradictions, in particular, need to be illuminated, explore, and discussed" (Fosnot, 1996, p. 29). Constructivist "designers" therefore rely on "instructional" strategies that will offer learners complex environments, assisting them in active and engaged exploration, helping them to construct and reflect on their own understandings (Karagiorgi & Symeou, 2005) and then to validate, through social negotiation, these new perspectives (Ertmer & Newby, 1993).

Fourth, knowledge (creation) is largely viewed as contextually bound: Knowledge "emerges" in contexts within which it is relevant. Consequently, advocates argue that the learning situation should be similar to the situation in which it is "applied" (Mandl, Kopp, & Dvorak, 2004) as they view transmission models to have a tendency to create "inert knowledge" (Whitehead, 1929), i.e. concepts that learners fail to use in problem-solving contexts other than the learning situation (or situations that are very similar). As pre-specified learning content and learning objectives are not congruent with constructivist views (Karagiorgi & Symeou, 2005), learner-centred approached to Instructional Design have also moved from the instructionist paradigm towards what has been called the problem solving paradigm (cf. e.g. Issing, 2002). From a problem solving perspective, environments should be created that are learner-centred, –directed and –controlled, offering "authentic" tasks, a plurality of perspectives and strategies (Gance, 2002). Constructivist pedagogies often include

hands-on, dialogic interaction with the learning environment. As an example, actually designing software would be preferred to simply being told how to design software (cf. Duffy & Cunningham, 1996).

Finally, from a (conservative) constructivist perspective, the traditional notion of assessment as an evaluation of knowledge is rejected due to the complete dismissal of objectivity, "correct" meaning, and "valid" judgement. However, assessment may be viewed as having the potential to render additional viable experiences for the parties involved. As such, constructivist pedagogies advocate a learner-centred, even individualized approach to assessment with self- and collaborative examinations and 'formative' evaluations that may contribute to the processes of creating knowledge. They also tend to focus on what in traditional terminology has been referred to as "authentic" tasks and "transfer", i.e. knowledge creation in situations that differ from the conditions of the initial learning environments.

Table 7

Differences between Cognitive Constructivist and Social Constructivist Perspectives

Aspect	Cognitive Constructivist	Social/Sociocultural Constructivist
The mind is located:	in the head	in the individual-in-social interaction
Learning is a process of:	active cognitive reorganization	acculturation/socialization into an established community of practice
Goal is to account for:	the social and cultural basis of personal experience	constitution of social and cultural processes by actively interpreting individuals
Theoretical attention is on:	individual psychological processes	social and cultural processes
Analysis of learning sees learning as:	cognitive self-organization, implicitly assuming that the learner is participating in cultural practices	acculturation, implicitly assuming an actively constructing learner
Focus of analyses:	building models of individual learners' conceptual reorganization and by analyses of their joint constitution of the local social situation of development	individual's participation in culturally organized practices and face-to-face interactions
In looking at tutored learning, we see:	an evolving micro-culture that is jointly constituted by the tutor and students	instantiation of the culturally organized practices of learning in higher education
In looking at a group, we stress:	the heterogeneity and eschew analyses that single out pre-given social and cultural practices	the homogeneity of members of established communities and to eschew analyses of qualitative differences

Note. adapted from Duffy & Cunningham (1996, p. 6) for higher education environments

While the framework outlined above serves as common ground for constructivist thinking, a wide range of pedagogies have been developed that place special emphasis on the various aspects of constructivist theory. Cobb (1994) distinguished between two major strands of constructivist perspectives – and thus, pedagogies –, namely, those that focus on

constructive activities of individuals and those that focus on collective knowledge construction (see Table 7 above). For the purposes of this study, pedagogies of the first kind of perspectives referred to as "cognitive constructivism" will be discussed in the following. The latter strand of "sociocultural" or "social" constructivist perspectives are subsumed under the "situative" perspectives heading below.

Various kinds of instructional, or, rather, *learning* technologies have been identified as generally being in line with cognitive constructivist theory (Cognition and Technology Group at Vanderbilt, 1996). It seems reasonable, however, to focus on two learning technologies providing environments that embody two pedagogical concepts brought about by an epistemological theory of (cognitive) constructivism, namely, *microworlds* (Papert, 1980) and exploratory or *discovery learning* (Bruner, 1961; Shulman & Keislar, 1966).

Microworlds are artificial environments in which learners can explore processes and rule systems while experimenting with different perspectives and combinations (cf. Schulmeister, 2002a, pp. 50-51). Here, technology environments provides mind-tools (Jonassen, 1996) for creative thinking, exploration, and knowledge construction (Papert, 1980) Koschmann (1996), for example, viewed the use of a computer programming language as a method to create original processes, procedures and structures and to actively explore and construct (or deconstruct) knowledge as a paradigmatic epitome for learning environments based on cognitive constructivism. He coined this paradigm *Logo-as-Latin*, thereby referring to the employment of the LOGO programming language used by children to explore mathematics and geometry (Papert, 1980). Today, however, Papert's (1980) microworld approach – and with it the *Logo-as-Latin*-paradigm – is often thought of as a cognitivist variant of computer-supported instruction, on account of (a) the rather instructionist notion that pre-constructed knowledge being embedded or 'hidden' in the microworld is to be discovered or retrieved by the learner (cf. e.g. Schulmeister, 2002a, p. 71) and (b) the *Logo-as-Latin* research approaches being so closely related to traditional behaviourist and cognitivist research (Koschmann, 2001).

While some have tried to use methods of combining "low-fidelity" text presentation and inter-personal discussion(e.g. Bellefeuille, Martin, & Buck, 2005), exploratory contexts can also be created by video and multimedia environments (Cognition and Technology Group at Vanderbilt, 1996). Here, technology-based environments are designed to present holistic, ill-structured, and relevant problem situations, often embedded in stories, games, or situational presentations. The "anchors" are required to be intrinsically interesting, problem-oriented, and

challenging (cf. Reeves, 1997) as well as sufficiently complex, meaningful, and authentic to preclude the construction of inert knowledge. Accompanying technology-based resources (e.g. video-DVDs or computer simulations) can be used by learners to explore possible contingencies as they try to resolve the problem. This approach, using contextual "anchors" for learners to explore and construct knowledge, has been called "anchored instruction" (Bransford, Sherwood, Hasselbring, Kinzer, & Williams, 1990; Cognition and Technology Group at Vanderbilt, 1992).

However, despite the fact that these approaches to cognitive constructivism generally acknowledge that learning processes cannot be separated from the very environment in which they take place (Mandl, Gruber, & Renkl, 2002), their main emphasis is on the relationship between the nature of the learning task in educational or training environments and its characteristics when situated in real use (Mayes & de Freitas, 2004). Constructivists here focus on providing for learning environments that facilitate the production, capture, storage, manipulation, and the distribution of created knowledge to the learning community (Li, 2001). As such, these constructivist pedagogies view "situativity" predominantly from a psychological perspective in which the main stimulus for learning is individual cognitive "conflict" or "puzzlement", or a "'deeply felt' (Doll, 1993, p.83) disequilibrium" (Gance, 2002, para. 5) - which is authentic to the social context in which the skills or knowledge are normally embedded (cf. Barab & Duffy, 1999).

Nearly every single aspect of constructivist pedagogies and its underlying axioms has been criticised (e.g. Fox, 2001; Kozloff, 1998; W. J. Matthews, 2003; Nola 1997; Phillips, 1995; Suchting, 1992; Solomon 1994; see also Gil-Pérez et al., 2002, for a recent overview). In fact, as Matthews (2000) observed,

> "Frequently the different aspects are treated as a package deal, whereby being a constructivist in learning theory is deemed to flow on to being a constructivist in all the other areas, and being a constructivist in pedagogy is deemed to imply a constructivist epistemology and educational theory" (Matthews, 2000, The Scope of Constructivism section, para. 10).

It is well beyond the scope and intent of this brief review to detail the various appraisals of cognitivist pedagogies; however, it seems appropriate to summarize five selected arguments that are of relevance to the subject of this thesis.

First, it follows from the working definition of socio-communicative competence delineated above that a competent handling of socio-communicative situations does, in fact, require a personal conception and appraisal process of the logic, dynamics, roles and tasks a situation demands and its requirements. Cognitive constructivist pedagogies therefore may have a valuable asset in their focus on experiential learning and cognitive concept development and a deepening of reflective understanding. In fact, cognitive constructivist pedagogies are essentially based on the general principle that "reflective abstraction is the driving force of learning" (Fosnot, 1996, p. 29). However, it may also be true that focusing solely on this aspect may preclude learners from paying sufficient attention to behavioural performance or applied skills.

Second, a related argument is that not for all subjects and knowledge domains, purely experiential approaches should dominate:

> "Some kinds of learners may feel more comfortable with facts, theories and the clarity of authoritative texts and received knowledge, at the very least as a starting point before they introduce their own opinions or attempt to solve problems." (Kalantzis & Cope, 2004, p. 51)

This may also be true in the social skills domain where the demand characteristics of unfamiliar and complex situations may often be overwhelming. Here, communication models and theory can serve as a means of facilitating meaning "taken-as-shared" (Cobb, 1991). A potential disrespect for behavioural or exemplary ("expert") models could also significantly slow down knowledge creation: As Bandura (1969) noted, observing models can "create" behaviour that had a zero probability of occurrence prior to the observation, even under conditions involving high motivation (see below). Following a related argumentative strain, Hess (2006) maintains that students, especially in costly educational systems, "may be geared towards maximum efficiency" (p. 110) and, thus, may expect instructors to help limit and pre-select sources of reference, to pre-structure learning content, and to exert guiding influence.

Moreover, if it is true that different people prefer different modes or 'styles' of experiential learning (D. A. Kolb, 1984, 1999), concentrating on experiential approaches alone may serve only a limited portion of all learners. This may also be the reason why even declaredly "cognitivist" approaches to computer-supported education do not forego teaching strategies traditionally utilized in instructional design, such as direct instruction (e.g. Bellefeuille et al., 2005).

Third, in contrast to instructional transmission models, cognitive constructivist pedagogies tend to place the individual learner and the learning environment rather than instructional processes at the centre of their endeavours. Often, this may imply an increase in the learners' personal responsibilities for learning processes for which self-regulatory competencies and motivation may be indispensable (Hipfl, 2003; cf. also Pintrich, Marx, & Boyle, 1993). Student's epistemological beliefs about the nature of knowledge and learning may also be influential predispositional factors potentially influencing the degree to which learners may benefit from constructivist learning environments (Schommer, 1993; Windschitl & Andre, 1998). These variables may therefore mediate the effectiveness of both the learner's personal conception of knowledge and skill gains and the negotiated consensus about an overall effectiveness of the implemented pedagogical approach.

Fourth, inextricably, there is a certain tension between constructivist philosophy as outlined above and the everyday functions and roles of instructors and advisors in higher education. The expectations of students, university programmes, instructors themselves, and, possibly, civil society organisations, may foster the perception of many that their teaching functions are not restricted to organizing and guiding their students' learning processes, but do encompass the preparation and provision of orientation, conceptual knowledge and specific subject matter, the design of learning environments pertaining to specific contents, and the assessment (and certification) of knowledge. Possibly, this friction may lead some to adopt some 'external' (first-order) observer's view. From this (less-than-radically-constructivist) perspective, learning progress may be re-defined in terms as the ability of learners to 'discover' what had been previously defined by the instructor and/or as the degree of development towards more complex (as opposed to 'simple') and context-dependent (as opposed to 'certain') knowledge, to be gained over time (as opposed to 'quick' knowledge), and as an increased subscription to the notion that learning can be learned (as opposed to 'innate ability' conceptualisations) (cf. Schommer, 1993; Windschitl & Andre, 1998).

And, finally, the claims that current technologies are inherently constructivist or may require a constructivist perspective cannot be sustained. Quite conversely, as one proponent of the constructivist perspective notes, their educational utilization often seems "retrograde in the sense that they largely incorporate behaviorist or information transfer strategies that are antithetical to a constructivist philosophy" (Gance, 2002, para. 2). For instance, interactive quizzes may require multiple choice responses inspired by the Mind-as-Container metaphor;

video-based vignettes may be transmissive or didactic in nature; the provision of hyperlinks is "customary in traditional instructional design, a reasonable and effective strategy, but not primarily constructivist in nature" (Gance, 2002, para. 14). By contrast, the coursewares and educational technologies available today often are deficient in providing elements of constructivist learning environments. For example, as Gance (2002) notes, learner-controlled hands-on interaction with the materials of the subject under study does not go beyond selecting among pre-determined choice points or examining predetermined content.

Situative Conceptualizations.

While cognitive constructivism focuses on the individual learner, his or her subjective experiences and attempts to generate meaning, social constructivists emphasize that the construction of knowledge or understanding evolves through social negotiation (Savery & Duffy, 1995). Social constructivists criticise conceptualisations of learning that focus exclusively on individual construction of knowledge as inadequate (Karagiorgi & Symeou, 2005, p. 18). Thus, theoretically, the same strand of critique applies to both instructional transmission models and the individual focus of cognitive cognitivism.

Yet, situative reconceptualisations differ in their various lines of thought as to what follows from the basic idea of "situated learning" that the cultural setting in which learning occurs influences the learning processes and outcomes. A varying and partially inconsistent array of relatively young and evolving positions and theories are therefore subsumed under this heading (Gerstenmaier & Mandl, 2001; Mandl, Kopp, & Dvorak, 2004; Wilson & Myers, 2000). As an extensive review is well beyond the scope of this chapter, it seems warranted to highlight distinctive concepts and views - notwithstanding that they may under-represent the range of different approaches to situative conceptualisations of learning.

While the relationship and the interaction are generally seen as preceding the individual (Gergen & Wortham, 2001), many concepts retain the traditional idea of individuals as being the carriers (or agents) of (their) knowledge and experiences. Examples that emphasize social and collaborative aspects of learning can be found in the sociocultural works of the Russian researcher Vygotsky (1978), Bandura's Social Cognitive Learning theory (Bandura, 1977; 1986, see below) and the recent writing of Bruner (1996). These approaches essentially emphasize social psychology traditions of viewing the "social" as an array of variables (cf. Burr, 1995), including the relationship between the individual and his or her social environment. As such,

they are not incompatible with the Mind-as-Container metaphor mentioned above (Bereiter, 2002; Bereiter & Scardamalia, 1996; cf. also Duffy & Cunningham, 1996).

Other situative-pragmatist conceptualisations reject the Mind-as-Container-metaphor oriented towards individuals and conceptualise knowledge, learning and learning processes altogether as "social constructions, expressed in actions of people interacting within communities" (Wilson & Myers, 2000, p. 59). Epistemologically speaking, these approaches attempt to overcome the body/mind-dualism of the aforementioned theories' riddles about how environmental (exogenic) and internal (endogenic) conditions of learning interconnect (cf. Gergen & Wortham, 2001, p. 123). They break with the traditional idea of personal agency so fundamentally embedded in Western thought (cf. e.g. Burr, 1995). Their unit of analysis is neither the individual nor the social and physical environment, but their interaction. Consequently, knowledge is not seen as an object, but it is constituted by and embodied in (a) people's relationships with other people and (b) their physical environment (which includes books, tools, artefacts, and other "learning objects") (Bereiter, 2002; Greeno et al., 1996). These two angles will be outlined further below.

On the one hand, learning activities and processes are seen as contingent and situated within the specific cultural (Lave & Wenger, 1991) or social (Rogoff, 1990) environments. Knowledge is also viewed as socially shared and created (Resnick, 1991). Thus, situative perspectives conceptualise learning as the strengthening of collaborative practices of the learning community and the participatory abilities and opportunities (cf. Greeno et al., 1996). Apprenticeship learning serves as an epitome for this model: Productive learning entails the legitimate participation of beginners in the community of practice - which first may be peripheral, but with increasing knowledge and proficiency, their participation becomes more central. Vygotsky's (1978) *zone of proximal development* (ZPD) is considered an important theoretical contribution. According to the ZPD model, 'experts' (or more capable peers), through collaboration and guidance – also called "scaffolding" – can help a learner reach developmental levels which he or she could not independently achieve. The use of the apprentice model to support learning in the cognitive domain has been termed "cognitive apprenticeship" by Collins, Brown, and Newman (1989) As a situative pedagogical concept, it "has gained respect and popularity throughout the 1990s and into the 2000s" (Dennen, 2004, p. 813).

On the other hand, situative approaches explain learning in terms of the relationship between learners and the properties of their learning environment. Separating learner, learning material, and context is seen as impossible, and, indeed, irrelevant, and as all learning is viewed as *in situ*, the very concepts of (task) "authenticity", "situativity", "meaningfulness", or "anchoring" become theoretically questionable (Young, Barab, & Garrett, 2000). Learning can be only considered in terms of the attunement of the learners to the offers ("affordances") and constrictions ("constraints") incorporated in the social and physical systems they interact with (Gerstenmaier & Mandl, 2001). In other words, learning is constituted when interaction improves between learners or between a learner and physical objects (including e.g. media) (cf. Mandl et al., 2004).

At first glance, both aspects discussed above do not seem incompatible with the working definition of socio-communicative competence which itself sees all situations of social communication as potential opportunities for development; cognitive apprenticeship in socio-communicative competence development may, for example, pertain to the scaffolding of personal construction of types of situations or the re-conceptualisation of behavioural actions as discourse activities.

However, As Wilson & Myers (2000) note, using situative approaches as a prescriptive basis for the design of learning environments somewhat contradicts their very idea: Communities of practice cannot be 'designed', but emerge within existing environments and constraints. Moreover, situated approaches basically contest the assumption that learning is a response to teaching. Thus, "situative" instructional models cannot resolve the inherent problem "… of how to design something that seems undesignable" (Wilson & Myers, 2000, p. 77). Nevertheless, technology systems have also been adopted by situationists, emphasizing its use as tools to foster collective knowledge building (cf. e.g. Young, 2003) (which requires interactive networks and feedback-loops typically insinuated by the contemporary buzz-word "web 2.0").

Koschmann (1996, 2001) sees computer-supported collaborative learning (CSCL) as an emerging paradigm in educational technology which is largely based on situative-pragmatist perspectives. CSCL systems support synchronous and asynchronous communication as well as face-to-face and remoted interactivity. One prominent CSCL example is Scardamalia & Bereiter's (1992) *Computer Supported Intentional Learning Environment* (cf. Schulmeister, 2002a, p. 82). Here, the cooperative construction of knowledge is supported through different

modes of communication (text, video, audio, animation) and the generation of "nodes", i.e. ideas or information related to the topic under study. A more recent example pertaining to the subject under study is *UniGame*, a web-based interactive platform used for collaborative simulation games targeted at fostering soft skills needed for reaching a consensus in joint problem-solving (Bouras et al., 2006; Pivec & Dziabenko, 2004).

Furthermore, some situative-pragmatist perspectives advocate evaluating the interaction the between the individual and the social and physical environment. Choosing this unit for analyses, however, would essentially restrict evaluation to its formative aspects. While some argue that summative or comparative educational research has not made essential contributions to educational practice (Schulmeister, 2002a), one could also argue that limiting educational research to formative analyses would essentially preclude effectiveness evaluations within the "curricular domain" (cf. Millman & Greene, 1989, see above).

Conclusions.

In summary, this review of perspectives on learning and instruction provides insight into factors that are critical for devising a developmental model of socio-communicative competence. It also demonstrates that, irrevocably and inescapably, such a model makes (a) fundamental assumptions about the nature of learning and instruction, (b) its evaluation and assessment, and (c) its starting points for technology integration. In the following, a model will be put forward and discussed in relation to the various perspectives outlined above. This developmental model is also to serve as a basis against which various features of this study can be referenced, as, for example, its instructional perspective and pedagogical approach, or the framework for evaluation and assessment.

DIGEST 2.3

In order to distil a developmental model of socio-communicative competence and starting points for supportive technologies, conceptualisations on instruction, learning, and assessment were reviewed under four paradigmatic headings. It was demonstrated that each set of perspectives makes fundamental assumptions about the nature of knowledge and skills, their development through learning and/or instruction, their evaluation and assessment, and offers different starting points for the integration of computer technologies.

2.4 A Framework Model of Socio-Communicative Competence Development

Based on the working definition socio-communicative competence and the review of perspectives on learning and instruction above, a developmental model for socio-communicative competence is outlined in the following. As it will be shown, this model guides instructional and educational practice, but also the evaluative research questions that are the subject of this study.

Components of an Integrative Theoretical Framework.

In accordance with the model of socio-communicative competence outlined above, its development in is situated and, thus, may necessitate exposure to situations of social communication which are novel to the learner. As we all encounter novel situations of social communication daily, it seems warranted to ask what separates "emergent" learning from socio-communicative competence development. One possible answer to this is threefold.

Firstly, as outlined above, socio-communicative competence here refers to professional types of situations. Thus, the concern here is neither with approaches to general life-skills development (e.g. buying food, asking for help, filing complaints) nor with therapeutic methods applied to psychosocial problems, specific mental or behavioural conditions (e.g. anxiety, lack of assertiveness, adjustment), or personality disorders that may affect social functioning.

Secondly, socio-communicative competence development is understood to represent a planned approach in which deliberately and intentionally targets at the respective knowledge, skills, attitudes and/or other characteristics of the participants thought to promote a conceptualisation or understanding of the demand characteristics of professional types of situations and/or their handling.

And, finally, as socio-communicative competence here refers to dispositional traits thought to limit behavioural variation across situations, rather than to specific, inconsistent states that would generate original situational handling "from scratch". Thus, everyday exposure to novel situations of social communication may not necessarily result in altered knowledge, skills, attitudes and/or other characteristics of a communicant.

Yet, what does teaching and learning psychology and pedagogy contribute to the question of socio-communicative competence development?

Interestingly, there exists one developmental meta-theory that pertains to both the relationship between social situations and learning on the one hand and the complex linkages between knowledge, skills, attitudes and performance on the other, namely, Social Cognitive Theory (SCT, also termed Social Learning Theory, SLT) (Bandura, 1977, 1986, 2001). According to SCT/SLT, there are two major processes in which learning occurs. These are termed *enactive* and *vicarious* learning.

Enactive Learning.

Originally rooted in cognitivist thinking and therefore utilising cognitivist terminology, SCT/SLT underscores the notion that learning – on the one hand – occurs through behavioural action (i.e. learning by doing) and environmental responses. In contrast to behavioural theories, however, the ingredient for learning here lies in the cognitive processing of the behavioural responses and environmental feedback information by a learner. Through abstraction, reflection, and rule generation, this information is transformed into "symbolic representations that serve as guides for action" (Bandura 1986, p. 51). In the case of experiential or *enactive* learning, this symbolisation process generates and what Bandura calls "generative conceptions" (i.e. schemata or procedural schemata or rules of action). Conceptualising this process in terms of the working definition of socio-communicative competence, enactive learning includes

(1) exposure to situations of social communication in which a communicant is challenged with

(2) *generating an understanding* of interaction processes and the demand characteristics of the situation.

(3) As the communicant "cannot not communicate" (Watzlawick et al., 1967, p. 49), he or she *acts*, i.*e. exhibits a (behavioural) pattern how to deal with or handle the social situation.*

(4) The communicant's conceptualisation of the interactional processes, the perceived demand characteristics of the situation, the perceived and sensed effects of the communicant's action ("intrinsic response") as well as the responses of the social environment ("extrinsic response", feed-back) can be subjected to (cognitive) analysis.

(5) The cognitive analyses may yield generative conceptions that provide (a) the rules for producing socio-communicative actions appropriate for the specific types of professional situations and (b) a system of reference to which situated socio-communicative actions can be compared (cf. also Kopp & Mandl, 2005).

(6) When subjected to cognitive analysis, the communicant may also gain a sense of mastery (or failure, respectively). While successes are thought to foster beliefs in the communicant's perceived self-efficacy, failures generally are thought to undermine it. Self-efficacy here is defined as the communicant's beliefs about his (or her) capabilities to produce designated levels of situated socio-communicative action that may affect the environment (cf. Bandura, 1994, 1997)

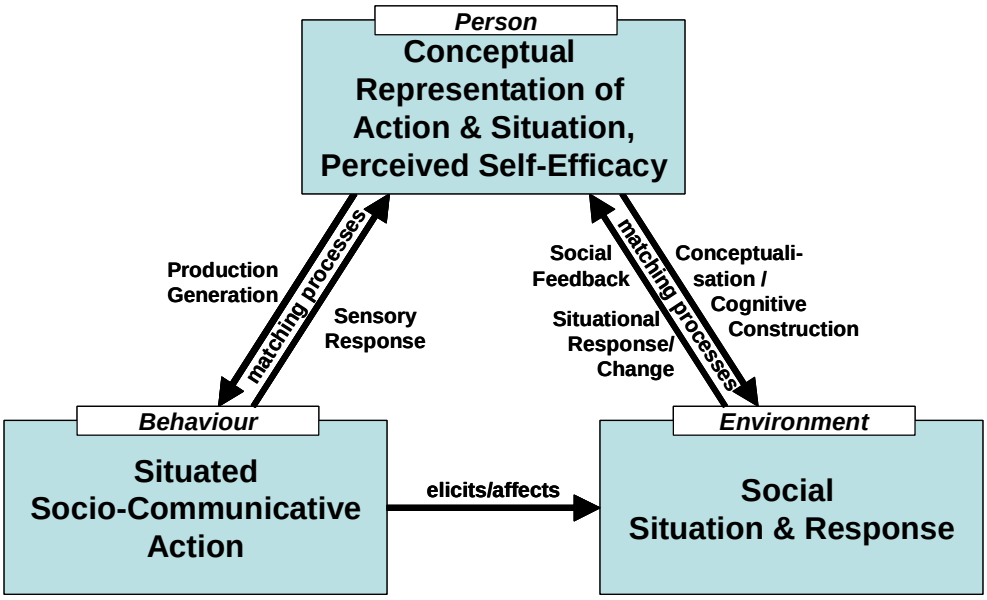


Figure 5. Processes of enactive learning of socio-communicative competence.

In summary, *enactive learning* as the combination of situated exposure to social communication and reflection processes is thought to generate the development of socio-communicative competence (see Figure 5).

The process described here as enactive learning has been described in similar terms by many educational theorists (cf. e. g. Ohlsson, 1996) as well as practitioners in the line of Dewey (1933, 1938), Lewin (1951) and D. A. Kolb (1984) who advocate experiential approaches ("learning by doing") to learning (cf. also e.g. Cantor, 1997; D. A. Kolb, Boyatzis, & Mainemelis, 1999; Kraft & Sakofs, 1988). According to Kolb's model (see Figure 6), immediate, concrete experiences form the basis for self-observations and reflections. These reflections are assimilated and transformed into abstract concepts, or, what Bandura (1986)

calls "generative conceptions". From these, novel implications for action can be drawn and actively tested. As such, they serve as guides for creating new experiences. Kolb's cyclical pattern underscores the notion that enactive learning is based on both production and reflection processes.

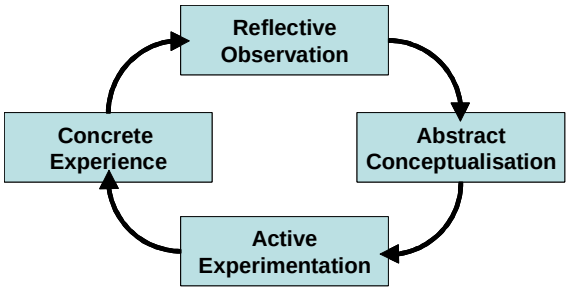


Figure 6. Cyclical pattern of learning (D. A. Kolb, 1984)

Enactive Learning also bears similarities and potential links with the *Reflective Practitioner* and *Theories of Action* models put forward by Schön (1983) and his colleagues (Argyris & Schön, 1987). For example, the individual's perceptions and conceptualisation of the communication process and situational demands can be thought to be guided by his or her *theories-in-use*. Similarly, reviewing the personal *theories-in-use* – either during the interactional encounter ("*reflection-in-action*") or subsequently ("*reflection-on-action*") – provide for the correction of inaccuracies and, thus, permit *single-* and *double-loop learning*. Whether a communicant's personal and largely implicit theory of socio-communicative action has become a *theory-in-use* or solely an "*espoused theory*" can only be observed during interaction. This also suggests, that in both theory and instructional practice, learning is not dependent on a specific mode of learning. For example, knowledge about communication models imparted in many communication trainings can influence perception and a communicant's conceptualisation of the interactive process. Likewise, knowledge about mediation stage models may shape a learner's judgement of and performance in mediation sessions.

Observational Learning.

On the other hand, SCT considers most human behaviour as being learned vicariously, i.e. in the absence of overt performance, through psychological matching processes referred to as *observational* learning though 'modelling':

> "Observers can acquire cognitive skills and new patterns of behavior by observing the performance of others. The learning may take varied forms, including new behavior patterns, judgemental standards, cognitive competencies, and generative rules for creating behaviors" (Bandura, 1986, p. 49).

Thus, the outcome of learning is not just described in simple, behavioural terms or the matching pattern between a modelled performance and the quality of behaviour as "imitated"

by the learner. In contrast to such a "response mimicry" conceptualisation, observational learning is, yet again, better thought of as a change in the learner's conceptions of "rules for generating variant forms of behavior to suit different purposes and circumstances" (Bandura, 1986, p. 48). As for socio-communicative competence, this may include a revision of pre-existing schemas for structuring the interactional processes and distinguishing the situational demands. It may also lead to more viable estimates of the attractiveness and functional value of modelled activities, and more precise judgements of their social efficacy.

On the personal level of the observer, changes in individual (propositional) knowledge, cognitive skills, but also attitudinal and emotional characteristics are thought to depend on two sub-processes mediating learning, namely,

- attention processes (e.g. the cognitive skills and prior knowledge of the observer, the complexity and discriminability of modelled activities, and their perceived functional value), and
- retention processes (e.g. modality, conciseness, and structure of symbolic transformation, i.e. the restructuring of information about external events; but also opportunities for cognitive rehearsal and cognitive enactment).

The perceived information about modelled actions and their embeddedness and applicability to types of social situations is thought to be constructively transformed into meaningful, albeit abstract symbolic representations (Bandura, 1986, p. 56). Again, as Figure 7 below suggests, reflective and feedback processes as, e.g. described by Schön (1983), likely enhance attention and retention. The cognitive representation of modelled actions can, for example, be tapped by thought probes (Bandura, 1986, p. 117). Correspondingly, the aforementioned Situational Judgement Tests also involve matching different potential courses of action to available symbolic conceptions.

However, the power of observational learning possibly lies in its potential to elicit behavioural actions that had a zero probability of occurrence prior to the observation, even under conditions involving high motivation (Bandura, 1969). Bandura (1986) suggests that whether the modelled actions will be performed by the observer and to what degree the performed actions match may depend on two other mediating sub-processes that refer to production and motivation.

- Production of behavioural actions may depend on physical capabilities and motor skills, but also on the availability and accuracy of conceptual mental representations as references and the possibilities to observe and compare the enacted performance to the reference system.
- Motivation to perform the modelled action may initially depend on the observers' expected consequences that are brought about should they themselves perform the observed actions. Modelling influences are also thought to increase the possibilities for behavioural actions when they *alter (dis-)inhibitions* over previously learned behaviour or *facilitate responses* in that they serve as social prompts creating motivational inducements for observers to behave accordingly (cf. Schunk, 2004).

  However, motivation also refers to intrinsic processes the most important of which may be personally set goals (e.g. Locke & Latham, 1990; Bandura, 1988), goal orientation (Dweck & Leggett, 1988; Ames & Archer, 1987) and self-efficacy beliefs of the learner, i.e. personal judgements of the learner's own capabilities to execute the respective behavioural actions and achieve an anticipated desired outcome. As Pintrich (2003) notes, "it has been a major finding from the earliest models of achievement motivation and behavior that when people expect to do well, they tend to try hard, persist, and perform better" (p. 671).

The various processes of observational learning are summarized in Figure 7. Vicarious observational learning can be conceptualised as a blend of perceptual processes of situated, modelled events and the cognitive transformation of these perceptions. It has been shown that observational can lead to novel patterns of thought and behaviour "which observers did not already possess, but which, following observation, they can produce in similar form" (Bandura, 1986, p. 49).

The observational learning process is not restricted to cognitive and behavioural responses. Various researchers suggest that through empathic vicarious experiences, modelling influences also elicit changes in emotional arousal and attitudes (e.g. Bandura & Rosenthal, 1966; Berger, 1962; Norton, Monin, Cooper, & Hogg, 2003). Thus, during the learning process, observers do not only "intellectually" take the role or perspective of the model (termed "cognitive empathy" by Gladstein, 1983; cit. in Dan & Hill, 1996), but possibly experience emotions similar to the model, a phenomenon termed "affective empathy" by

Gladstein (1983; cit. in Dan & Hill, 1996). Thus, observational learning is more than a mere imitation or response mimicry process.

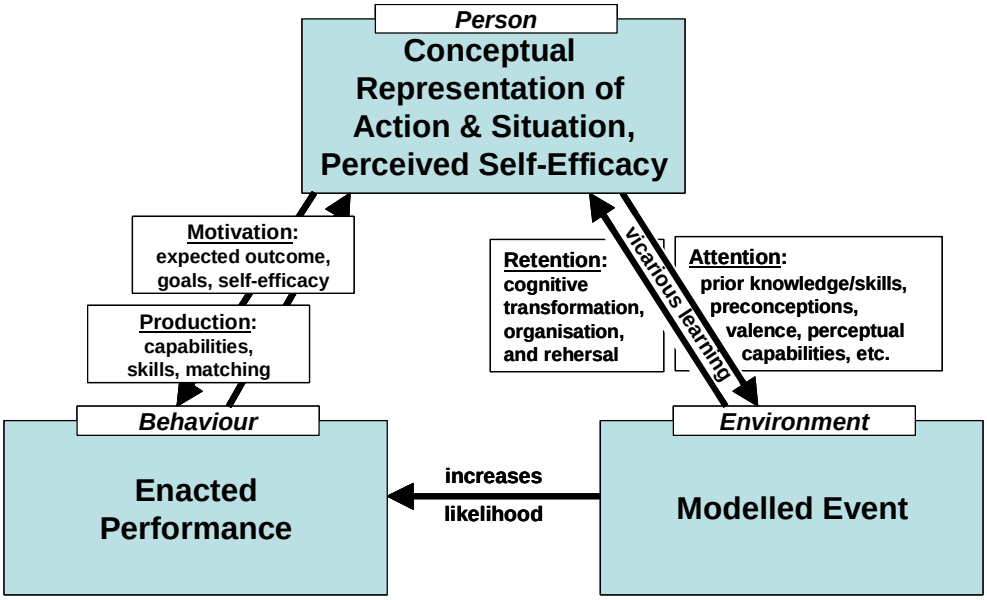


Figure 7. Processes of observational learning.

In summary, vicarious observational learning, mediated through processes of attention and reflection, impinges on cognitive and emotional potentials of the observer that, prospectively, can transformed into behavioural performance.

SCT/SLT as a Meta-Theoretical Base.

As Gibson (2004) noted in a recent review, SCT/SLT "is identified in the adult learning literature as one of five traditional theories of adult learning (Merriam & Caffarella, 1999) and as a meta-theory of learning for human resource development (HRD) (Swanson & Holton, 2001)" (p. 193). Albeit originally based on cognitivist theory and research, this meta-theoretical framework may also comprise links to constructivist and situative pedagogical approaches to learning and development. For example, it can be suggested that all ways of learning are thought to be triggered or motivated by deliberate or unplanned "perceived" inaccuracies or what constructivists would call "disequilibrium". Moreover, within the interactive, reciprocal relationship between the cognitive processes of the individual, the behaviours and the environment put forward by SCT/SLT, "perception" and "retention" can be appropriately thought of in terms of sub-processes of "mental construction". Gibson (2004) cites Torraco (2002) who related the basic tenets of SCT/SLT to help explain situative learning in Lave & Wenger's (1991) communities of practice concept. The essentially agentic perspective

maintained by the SCT/SLT framework (Bandura, 2001), however, as well as its prescriptive nature for instructional and evaluative purposes may preclude its employment in both "pure" constructivist and situative pedagogies.

Framework Integration and Links to the Praxis of Training.

In summary, the discussion of SCT/SLT and related theories above suggests that changes in an individual's cognitive-affective schemas or what Kolb (1984) calls "abstract concepts" used for both the judgement of a professional type of situation and the corresponding internal cognitive-affective appraisal as well as the generation and execution of behavioural responses can be brought about by

(1) imparting conceptual knowledge (as, for example, in theoretical communication models) that supports situational feature detection, schema generation and adaptation, or behaviour generation;

(2) vicarious observational learning as a preparatory stage for

(3) enactive learning or what Kolb (1984) calls "active experimentation" to improve a systematic behavioural variation and test the applicability within situational types and across situations (learning-by-doing); and

(4) solitary and collaborative reflection processes (cf. e.g. Schön, 1983) focussing on situational features, an individual's idiosyncratic appraisal (what Kolb calls "concrete experiences"), and possible alternatives to the communicant's behaviour.

These four elements or "paths" of the framework are depicted in Figure 8 below.

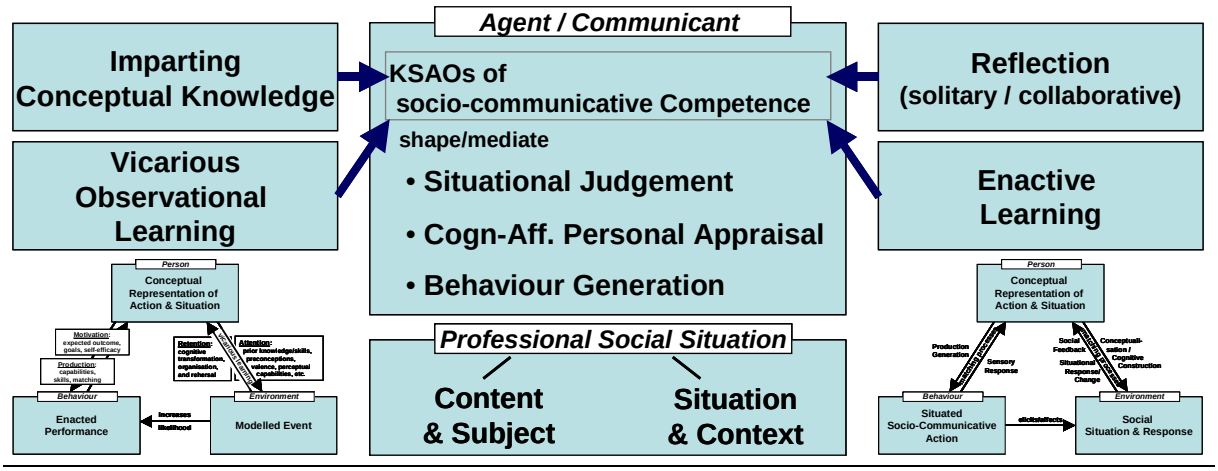


Figure 8. Four Paths to Professional Socio-Communicative Competence Development.

Having identified these four elements, a few words shall be dedicated to describe how these elements are typically implemented in the practice of social skills training. While the above mentioned "paths" may sound somewhat theoretical, they are closely linked to most practical educational activities and training practices. In turn, this may help practitioners to relate their knowledge and experience to the framework elements.

Imparting Conceptual Knowledge Skill-oriented training programs often begin with the provision of a description of skills or behaviours to be learned. This may simply include a presentation of learning points (e.g. "How to listen actively: …") or more general principles underlying target performances ("Summarize the statement concisely in your own words before you reply.") which have been termed "rule codes" (Decker, 1980). In a more complex form, knowledge about a more theoretical concept is imparted. This may be an accumulation of rule codes to describe the essence of broader or long-term strategies or attitudes as, for example, in scales used for client-centred therapy training purposes (Sachse, 1992; Tausch, 1973; Truax & Carkhuff, 1967). It may also be in form of explanatory models intended to enhance situational understanding such as the *Parent-Adult-Child-model* used in *Transactional Analysis* (Berne, 1964) or Schulz von Thun's *Square of Communication* (Schulz von Thun, 1981, 2004b), or a more strategic action model may be presented such as a mediation process stage model (Moore, 2003; Redlich, 2004a).

Vicarious Observational Learning To illustrate both target performances and the more complex models in training, small model sequences are often offered to trainees to enhance understanding of the knowledge imparted. Sometimes, video sequences are used to exemplify target behaviours or small, role-played live sequences are used demonstrate processes and impact. Trainees observing the models presented are thought to symbolically code the observed behaviours and possibly engage in a symbolic rehearsal of later reproduction. Here, contrasting positive models showing the desired behaviours with such who present what is not correct can enhance transfer (Baldwin, 1992; Taylor, Russ-Eft, & Chan, 2005).

Enactive Learning A major component of social skills training is enacted or experiential learning. Trainees are encouraged to practice the target behaviours in (and also outside) the classroom in different situations and cases. Often, in-class live situations are evoked or role-play simulations are used. In their meta-analytic recent review of skills training studies, P. J.

Taylor at al. (2005) found that, in line with the general expectation, the higher the amount of behavioural practice (as measured by the number of training hours) is positively associated with larger training effects on the development of procedural knowledge-skills. They also found that the use of work-related scenarios developed by trainees themselves may generally lead to better transfer-when compared to pre-devised scenarios provided by trainers.

Reflection (solitary / collaborative)

In social skills training, in-class practice regularly comes along with instant feed-back provided by either the interaction itself or by solicited feedback from either the live counterpart or the role-play partners. Feedback may also be provided by the trainer during or after the enactive phase. A review of learning points, rule codes, a group-based discussion of situational types in which the behaviours displayed were advantageous or possibly detrimental are also often used in training. All of these measures are intended to (a) increase target skill performance, (b) improve retention of knowledge and skills, (c) enhance transfer, and (d) they are also thought to be connected to motivation and self-efficacy.

DIGEST 2.4

In line with the working definition, the judgement of a professional type of situation, the corresponding cognitive-affective appraisal as well as the generation and execution of behavioural responses are thought to be developed by imparting conceptual knowledge, vicarious observational learning as a preparatory stage for enactive learning and subsequent solitary and collaborative reflection processes.

2.5 Computer-Supported Social Learning (CSSL)

In the foregoing review of perspectives on learning and instruction, paradigmatic approaches to computer integration were detailed. In summary, it can be concluded that computer-based collaborative learning (CSCL) employs social learning theories as a theoretical foundation. However, when individual computer-based or –supported learning rather than CSCL is considered, which meaningful starting points for technology integration or "blended learning" arrangements can be distinguished?

Defining Computer-Supported Learning.

As indicated above, the term "blended learning" currently enjoys widespread use (Oliver & Trigwell, 2005). Recent reviews (Driscoll, 2002; Oliver & Trigwell, 2005) note that the term is "ill-defined" (Oliver & Trigwell, 2005, p. 17) and used to denote different concepts or phenomena, however. The use of the term to denote the combination of computer-based preparation and subsequent participation in face-to-face training courses would therefore leave something to be desired. For example, in line with the common use in research (Oliver & Trigwell, 2005) it seems to refer only to the combination of different *training delivery formats* (e.g. 'traditional' face-to-face and technology-based instruction) as indicated above. However, critics would argue that most, if not all, learning activities involve technologies of some kind, even face-to-face training. And, vice-versa, the inverse may also apply in contemporary higher education environments; most, if not all, purposive technology-based or –supported learning activities may invariably entail some face-to-face interaction within classroom settings (cf. Draper, 1997).

Reverting to terms such as "computer-based" or "web-based" delivery provides for a conceptually less ambiguous stance. However, these two terms are often used as contradicting expressions for technological potentials in physical delivery (e.g. local CD-ROM vs. internet/online delivery, bandwidths, access speed, revisability, shareability, collectivity, etc.). They also seem to refer to different historical 'paradigms' of learning and instructional technology, with (behaviourist) computer-aided (programmed) instruction on the one hand (e.g. Koschmann, 1996, 2001) and an exploratory, web-based (constructivist) learning environment on the other – despite the fact that both "computer-based" and "web-based" delivery formats can potentially distribute the very same content and allow for the same level of interaction.

In view of this fact it seems appropriate for a definition to make explicit reference to the very medium technologically required for the purposes of delivery, namely, the computer. This is not meant to insinuate that the computer is to be distinguished from other means of delivery by virtue of its specifics, but, rather, to let it represent the ways, channels, and possibilities of the digital processing and interaction typically covered (but not exclusively featured) by this medium. As such, the term explicitly includes appliances available today and in the future (e.g. handhelds, personal desktop assistants/PDAs, smart cellular phones, etc.) that allow for the same form of delivery and interactive processes.

In the following, this idea will be referred to (in the headline of this thesis and below) the by use of the term "computer-supported" (learning or instruction), as the viable alternative - "computer-mediated" - lacks discriminatory power and could imply that the various media used (e.g. books, software, lectures) have fixed or essential qualities (cf. Oliver & Trigwell, 2005).

The 3C Model.

In their so-called "*3C-model*", Kerres and de Witt (2003) suggested that blended learning arrangements differ in their design and embodiment of three didactical component categories, namely, 'content', 'communication' and 'construction'. '*Content*' refers to the learning materials being made available to the learner. Interpersonal exchanges (face-to-face or virtual, analogue or digital, synchronous or asynchronous; peer-to-peer, learner-to-tutor and tutor-to- group and vice-versa) is subsumed under the '*communication*' heading. And, finally, the '*construction*' component refers to all didactical elements that are intended to assist and guide both individual and cooperative learning activities (e.g. tasks or assignments), varying in their degree of complexity (from multiple-choice to problem based learning).

Kerres and de Witt also suggest that these three component categories are given different relative weights by blended learning delivery designers. This means that different blended learning arrangements vary in the amount of time a learner engages in the absorption of informational content, in communicative activities, and the solving of constructive tasks. This principle goes so far, as Kerres and de Witt point out, that neither component can be considered a necessary element for blended learning arrangements.

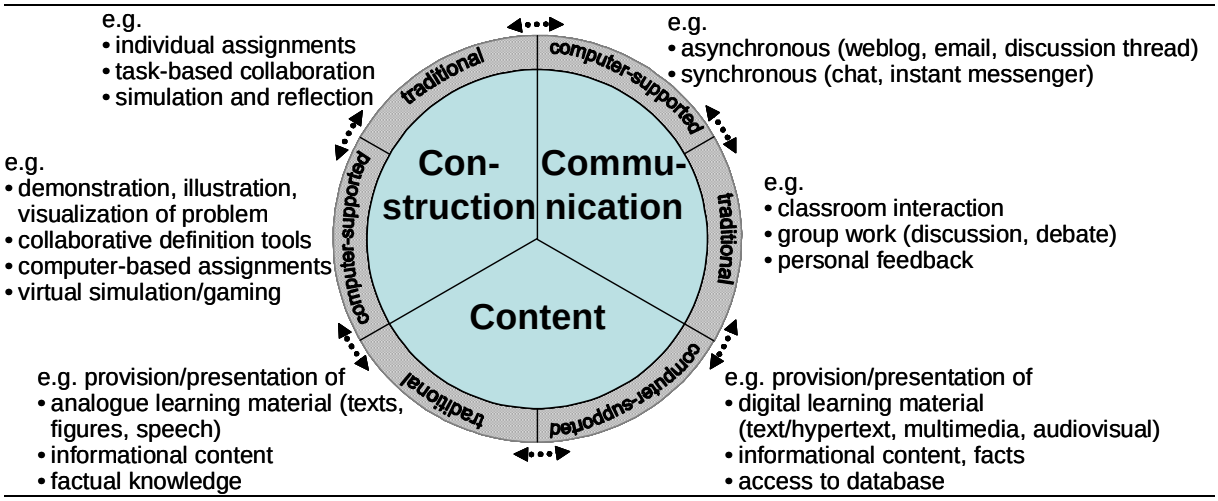


Figure 9. Extended 3C model based on the original model proposed by Kerres and de Witt (2003).

For practical purposes, however, this model warrants an extension as the effectiveness of the respective delivery modes is only being discussed but not explicitly taken into account by Kerres and de Witt (2003). To be able to understand the design and assess the probable effectiveness of a blended learning arrangement, one must not just know the relative weights of the three components as a whole, but the design and relative weights of the three components in the various delivery modes and their interaction, as depicted in Figure 9 above.

Computer-Supported Social Learning (CSSL).

The integrative framework of socio-communicative competence development will now be expanded to incorporate pedagogically meaningful starting points for individual computer-supported social learning (CSSL) as a complement or supplement to traditional forms of instructional delivery. These forms will be briefly reviewed first, drawing on the basic *3C model*, to establish the distinct strengths of computer-supported learning within the framework in order to generate an intelligent "blended learning" mix.

Probably the most traditional forms of instructional delivery are books and lectures. The German expression for "lecture", *Vorlesung*, has its etymological root in *lesen* (to read), and is combined with the prefix *vor-* (literally: 'before'or 'in front of' others); the expression thus still denotes the delivery form's instructional-transmissional character. Speaking in *3C* terms, the *content* component is predominant. The extended 3C model also suggests that, if construction and communication components are scarcely utilized in this traditional form of classroom delivery, it may be harmlessly – or even advantageously – replaced by computer-supported forms of delivery.

As maintained above, conceptual knowledge relating to socio-communicative competence may be imparted by means of books and lectures; however, lasting changes in the reader or listener will be observable only when this conceptual knowledge is appropriately applied. To this end, didactical elements that contribute to knowledge construction are indispensible. They are needed to make feasible (a) a novel judgement of an external situation which differs from the situations used for learning purposes (but may be similar in type), to make way for (b) an altered internal cognitive-affective appraisal of a situation and/or to improve (c) generation and execution of behavioural responses. For this to happen, didactical elements that fall under the '*communication*' and '*construction*' component headings in the *3C model* are essential.

Opportunities for providing models are essentially restricted to low-fidelity descriptions or case studies in text books and a limited number and variety of methods that can be used in traditional lectures. By contrast, story books or novels may contain high-fidelity descriptions and afford conceptual changes as they model people's experiences in a directly accessible and personalisable way. This may be one of the reasons why narratives are a primary medium for imparting conceptual knowledge in technology supported professional training (cf. C. E. Hansen & Williams, 2003; Jonassen, Strobel, & Gottdenker, 2005).

However, traditional text books do not and lectures scarcely provide for either enactive learning opportunities (which, in fact, can be considered one *via regia* of individual 'construction' in the *3C model*) or collaborative reflection processes (which, in terms of *3C*, consist of didactical elements subsumed under the 'communication' and 'construction' headings). In fact, the latter are viable options in face-to-face small-group training courses or tutorials, however. Here, as noted in the introductory chapter of this thesis, simulation and role-play are practical methods that offer chances for exposure to specific types of professional socio-communicative situations, for observational learning and solicited feed-back. As noted above, books, lectures, and small-group tutorials are traditional methods of instructional delivery used extensively in mediation training programmes. The question remains, however, as to what extent these didactical elements supporting knowledge construction and communication processes can be fruitfully supplemented by support of the computer.

As professional socio-communicative competence refers to specific types of professional situations, its conceptual models and its development can be expected to pertain to these situational types. It is exactly here where traditional forms of delivery opting for one of

the four developmental ways detailed above may benefit from the presentation of complex socio-communicative situations by means of digital media (video and/or audio scenes). As Bandura noted, media technologies generate "major changes in the models of behavior to which people … have access" (Bandura 1986, p. 55). Television (Seels, Fullerton, Berry, & Horn, 2004) and videodisc (Fletcher, 1991; Schroeder, 1985; Schroeder et al., 1984) have been found to be helpful instructional media.

Preferably, the multimedia scenes are embedded in a storyline or narrative context to ease encoding of conceptual knowledge (cf. Jonassen et al., 2005). Renkl and Atkinson (1992) suggested that skill acquisition is assisted by presenting problems with high-fidelity, worked-out examples that essentially capture the problem formulation and its solution. They argued that computer technology may support the critical variable in this learning process, namely, assisting the learner to explain the exemplary solutions to themselves. An excellent example is found in Gentry's (1992) early description of uses of computer aided technology for mediation trainees. In her sequential instructional programme, learners were

(1) acquainted with an introductory analysis of mediator tasks and roles (this relates to the submission or *imparting of conceptual knowledge* and schema generation) before they were
(2) asked (by means of provided checklists) to identify mediator roles and tasks in a series of video scenes (this relates to *situational feature detection* as well as processes of *observational learning*); learners were then
(3) asked to compare their own analysis with that of a professional expert (this relates to *solitary reflection* for schema adaptation).

These three processes can be understood to deepen conceptual knowledge. When situations are presented and elaborated on in such a way, these situations as well as the behavioural options may be conceptualised as serving as modelled contextual "anchors" or "cues" in critical or typical professional situations. On the one hand, task-oriented responses my be directly elicited by such situational cues (Henninger, Mandl, & Hörfurter, 2003). On the other hand, learners typically profit from reflection-through-comparison in that the quality of their reflective analyses is enhanced. The reflection processes induced by this model are of the "reflection-in action" and "reflection-on-action" types discussed above. However, as Scheepers and Nuldén (2000) point out, these two types rather relate to past experiences, while professional action also requires thinking about potential applications, i.e. how to perform and apply the knowledge gained by this model in future. Citing Cowan (1998), Scheepers and Nuldén (2000) refer to this type of learning as "reflection-for-action" (see also McAlpine,

Weston, Beauchamp, Wiseman, & Beauchamp, 1999, for a related discussion). When novel situated problems (or cases) are encountered, the communicant may revert to past cases (i.e. modelled solutions or reflection analyses) from memory and draw on (narratively) encoded conceptual knowledge to resolve the problem, as case-based reasoning theory (Aamodt & Plaza, 1996; Kolodner, Owensby, & Guzdial, 2004) suggests.

In sum, this learning sequence, by use of a narrative and/or situational anchors and by prompts for reflection, provides an excellent add-on to books and lectures traditional systems of instructional (or conceptual learning) delivery.

As outlined above, learning processes as conceptualized in terms of Social Learning Theory can be supported by means of computer technology. However, computer-supported social learning is not limited to (the nevertheless powerful) aforementioned processes, as will be outlined below. After having completed the learning sequence of Gentry's programme detailed above, learners were

(4) asked to identify with the mediator in a series of mediation scenes presenting a specific demand characteristic in the on-going inter-personal interaction. (This again relates to processes of *situational feature detection*.)

(5) The viewers were then prompted to formulate what they would say and do as the mediating person as a response to what happened in the scene. (This is an idiosyncratic element of video-based interactive courseware and relates to a special kind of behavioural generation or *active experimentation*.)

(6) Learners were subsequently given the opportunity to view the responses of the actual mediators acting in the scene. (This again relates to processes of *observational learning*.)

(7) Learners were then given the task to identify roles and tasks of the actual mediator (and thus, to *reflect* on *situational features* and *generative concepts*) before

(8) they were again asked to compare their own analysis with that of a professional provided by the programme (this again relates to *solitary reflection* for schema adaptation).

This latter series of processes can, again, be understood to deepen conceptual generative knowledge and, again, to provide contextual anchors or "cues". In addition, however, the interactive element of this video-based courseware allows for a specific kind of learning process, namely, a form of experimental behavioural response to a realistic, complex situation and task. Thus, this approach provides an opportunity for a crude form of enactive learning and, thus, allows for experiential learning and much more elaborate reflective processes. Even without the provision of feed-back, such a task-oriented approach with "'triggering' and guiding questions" has been found to be "most beneficial for the learning

process (Cennamo, 1994)" (Bakx et al., 2002, p. 344). Presumably due to the more developed level of generative concepts, reactions towards such an approach tend to be more positive in somewhat advanced learners (Bakx et al., 2002). A potentially mediating factor is the students' level of familiarity with computers and computer learning. Supplementing the earlier view that computer literacy may be a necessary prerequisite for effective knowledge acquisition, recent research also suggests that higher levels of familiarity with computers may indeed hamper inquiry learning (Wecker, Kohnle, & Fischer, 2007).

In sum, this approach may serve as a viable step and link between the deepening of conceptual or generative knowledge and the actual performing: As a recent study (Hollandsworth, 2005) suggests, such an approach to a computer-supported exploration of socio-communicative situations and behavioural options may reduce the high cognitive load often present in face-to-face delivery of interpersonal training (cf. van Merriënboer, Jelsma, & Paas, 2002). Moreover, inasmuch as this process reflects on specific socio-communicative (or, as in this case, mediator) tasks and the specific conditions of task attainment, it can also be understood as facilitating the building of implementation intentions (Gollwitzer, 1999). Thus, advanced learning technologies may not just

In summary, when complex socio-communicative situations are presented by means of video or audio, learners may be asked to

- detect situational features, for example, by applying conceptual knowledge imparted previously or concurrently (as in books or lectures or through courseware);
- reflect on their individual cognitive-affective appraisal of the situation presented;
- generate a (written) response in line the situational demands and either their own cognitive-affective appraisal or set goals and motives of the actor;
- observe, evaluate and rate various behavioural options for communicative action presented subsequently;
- reflect on feed-back offered after having rated the list of alternatives; and/or
- reflect on their own generated response by comparing it to behavioural options or expert analysis presented.

Computerized analyses and feed-back loops of the learner's inputs would seem helpful to specifically target the individual's conceptual or even attitudinal (self-efficacy, motivation) learning processes (cf. Mory, 2004; Musch, 1999); however, due to limited resources and the lack of sophisticated technology, feed-back offered by computer-based training programmes is presently restricted to the options presented to and evaluated by the learner.

In sum, therefore, soft-, course- or hyperware that contains tasks such as those listed above can be hypothesized to be constructively employed

- after learners are acquainted with basic conceptual knowledge that helps identifying situational features that are crucial to generate effective responses; and
- before learners are exposed to experiential or behavioural training.

The overall approach to employing computer technologies, focusing on individual learners, will be referred to as Computer-Supported Social Learning (CSSL) below. For purposes of comparison, a (non-exhaustive) list is detailed in Table 8, characterizing the traditional approach to delivery of conceptual learning (i.e. books or lectures), Computer-Supported Social Learning, and planned behavioural skills development by use of face-to-face training (CSSL).

In line with the working definition of socio-communicative competence and the developmental framework outlined above, CSSL should positively affect learning (and/or instructional) efficacy. In the following, relevant research will be briefly reviewed to generate a list of potential effects or outcomes.

Table 8

Comparison of Conceptual Delivery, CSSL Courseware, and Face-to-Face Training

Written Materials / Lectures	Interactive CSSL Courseware	Face-to-Face Training/Coaching
Description and presentation of information, models, and theories	Depiction of situations, behaviours, and implementation	Guided exploration and experience of behaviours and encounters
(Abstract) concepts as focus	Typical critical incidents (of roles, situations, behaviours) as focus,	Critical incidents as 'induced' or experienced by learner as focus
Stimulating understanding	Stimulating application of concepts, situational judgement, and reflection	Stimulating action, observation, and reflection-in/on-action
Re-readable written materials / Non-Repeatable explanation in lectures	Repeatable viewing and reflection on situations or behaviours shown	Non-repeatable, idiosyncratic situations, behaviours, and incidents as basis for reflection
Content adapted to subject matter and target audience	Content and Interactivity tailored to learning goals of target group	Contents and Setting negotiated by coach/instructor and learners
Limited number of case study presentations feasible	Great number of field examples and situations feasible	Simulation or exploration of limited number of settings
Limited variability in stimuli (text description / media presentation)	Multi-media, multi-modal, high-fidelity stimuli feasible	'Full immersion' feasible
Application errors remain unrecognized unless explicitly described by instructor or unintentionally by learner	Planned recognition of critical errors or incidents by learners	Limited anticipatability of occurrence and recognition of errors or incidents
No correction of misunderstandings beyond "Questions and Answers"	Correction of misunderstandings by way of elaboration and indirect feed-back / unplanned recognition by learner	Correction of misunderstandings by way of direct feed-back
Focus on teaching	Focus on interaction	Focus on training/coaching
Pace selected by reader / lecturer	Pace selected by learner	Pace negotiated by coach/instructor and learners
Motivation of learner / attendee, literacy, and knowledge levels main conditions for learning	Motivation of learner / attendee, and computer literacy main conditions for learning	Motivation and relationship between stakeholders main conditions for learning
Accessibility and processing depth influence progress	Usage level, quality of models, and processing depth influence progress	Learning culture (e.g. error-friendly environment) and feedback skills of stakeholders and influence progress
Typically no quality control beyond sales figures / evaluation results; content revisions typically due to advances of theory/concepts	Content, feedback, leaner's reactions and perceived quality controllable and easy to adapt/improve	Assessment and improvement of training and feedback skills complex

DIGEST 2.5

Computer-based delivery technology may not just replace traditional forms of delivery, such as textbook presentation and lecturing. It may also provide high-fidelity solutions to support individual knowledge construction processes required for social learning, for example in the form of multimedia-based presentation of socio-communicative situations that serve as prompts for interactive -exercises and tasks. Even without reverting to computer-aided collaboration, computer-based solutions of this kind can be hypothesized to be constructively employed, for example, after learners are acquainted with conceptual knowledge and before learners are exposed to experiential communication training.

2.6 Potential Outcomes of Computer-Supported Social Learning

It was noted in the introductory chapter that despite the general lack of research aiming at evaluating the efficacy or effectiveness of blended learning solutions for soft skill development – especially when compared to traditional delivery modes such as books, role-play, and video and multimedia presentations – there are a number of studies attempting to evaluate the efficacy or effectiveness of employing computer technology for social skill development. Excluding parameters of formative evaluations, Table 9 provides an overview of outcome parameters used in exemplary research studies for which significant results had been found.

In general, the outcome parameters found can be classified under three broad headings: (1) *performance* (e.g. grades, knowledge tests, or skill assessment), (2) *attitudes towards self* (e.g. self-efficacy, self-confidence, self-rated abilities), and (3) *attitudes towards technology-supported learning* (e.g. motivation, satisfaction levels, or course ratings). This finding is much in line with earlier reviews (Chumley-Jones, Dobbie, & Alford, 2002; Phipps & Merisotis, 1999).

Performance.

Most multimedia learning evaluation studies revert to knowledge tests to measure knowledge gains; only few use more elaborate or complex assessment instruments such as performance-based assessments or rated essays (C. E. Hansen & Williams, 2003). By contrast, as concluded above, the complex nature of socio-communicative action calls for instruments measuring skills beyond knowledge; in fact; many studies of technology–based or –supported learning for social skills listed in Table 9 below employ situational judgement tests and/or performance-based assessment strategies.

Table 9

Evaluation Outcome Parameters in Studies of Technology–based/–supported Learning for Socio-Communicative Skills

Authors	Study Description	Outcome Parameters/Instruments
Bildat (2000) / (2003) Bielecke (2005)	Computer-based training programme for critical communication skills (Quasi-Control/Post-test only)	• Situational Judgement Test
Cauble & Thurston (2002)	Computer-based training programme for social work practice education (Pre/Post and Quasi-Control)	• Knowledge Test • Competency (Self-efficacy) Rating Scale (self-rated)
Evans, Petrakis, & Swain (2001)	Web-based multimedia exercises for virtual social work placement (Implementation study)	• Attitudes towards programme
Fleetwood et al.(2000)	Web-based interactive multimedia exercises for case-based ethical communication to patients (Post-test only, Random Control)	• Knowledge Tests • Performance (standardized patients) • Self-Confidence Rating • Satisfaction with WBT/overall course
F. C. B. Hansen et al. (1998)	Computer-based training programme for paraphrasing and perception checking (Pre/Post)	• Knowledge Test • Self-reported confidence in listening skills
Gentry (1992)	Computer-based video training for mediation trainees (Post-test only)	• Attitudes towards CBT/ Satisfaction with CBT • Self-reported learning gains
Holsbrink-Engels (2003)	Computer-based simulation training for telling bad news (Pre/Post)	• Performance in computer-based assessment (rated) • Knowledge Test (naming essential characteristics of the skill) • Classification Test (ability to identify bad news situations)
Irvine, Ary, & Burgeois (2003)	Computer-based training programme for communication skills in professional care-giving (Randomized controlled trial video-based lecture vs. CBT format)	• Situational Judgement Tests (behavioural intent / knowledge) • Self-efficacy (Confidence in performing best-choice option)
Kilburg (2005)	Computer-based training programme for the facilitation of group-based problem solving (Pre/Post)	• Knowledge Test
Poulin & Walter (1990)	Computer-based training programme for interviewing skills in social work (Pre/Post)	• Attitudes towards programme
Wiecha, Gramling, Joachim, & Vanderschmidt (2003)	Web-based observational and reflective learning of medical interviewing (Pre/Post)	• Self-reported level of understanding of in concepts of medical interviewing
Williams et al. (2001)	Computer-based training programme for psychotherapeutic skills (Randomized controlled trial lecture vs. CBT format)	• Self-reported learning gains (lecture > CBT) • Performance-based practical skill assessment (lecture < CBT)

Post-Training Attitudes.

Self-efficacy (Bandura, 1982, 1986, 1997, 2001) in this context pertains to the beliefs of learners as to their capabilities to handle particular situations in certain ways or at designated

levels of performance. These beliefs are thought to be adaptive in nature and influenced by the learner's previous attainments (or accomplishment failures), but also processes of observational learning with models performing tasks similar in kind (Bandura, 1982). Due to the relationship repeatedly found between self-efficacy beliefs and performance (Alvarez, Salas, & Garofano, 2004; Stajkovic & Luthans, 1998), it seems reasonable to use self-efficacy beliefs as an outcome parameter. However, self-efficacy beliefs are also an important factor mediating learning through motivational processes. Incentives or the motivation to learn may be dependent on neither overly optimistic nor pessimistic of self-efficacy expectations (Bandura, 1997). As Pintrich (2003) put it,

> "On the one hand, from a motivational perspective, it would seem that having as high a possible efficacy and competence beliefs would be useful and keep students motivated. On the other hand, from a self-regulatory perspective, if students are consistently overestimating their capabilities, they might not be motivated to change their behaviour in the face of feedback that provides them with information about their weaknesses" (p. 671)

It may be assumed, therefore, that educational or learning systems should help learners to arrive not at maximal but optimal levels of self-efficacy. However, there is a general lack of research pertaining to the issues of how to "calibrate" self-efficacy beliefs and how they are linked to performance (Pintrich, 2003; Stone, 2000). For example, in their randomized controlled trial, Williams et al.'s (2001; cit. in Berger, 2004) compared their computer-based training (CBT) programme for complex social skills in psychotherapeutic assessment to a lecture format using equal video contents and duration. Results indicate that learners in the CBT condition rated themselves as having learned significantly less on most of the subjective ratings of perceived ability than the lecture attendees. Interestingly, however, the ratings of practical skills in a subsequent performance-based assessment (mental-state examination) were significantly higher in CBT learners. Possibly, an active exploration and reflection on professional (or expert) behavioural demonstrations may augment the "standard" or frame of reference against which one's own expected performance is compared (e.g. Chumley-Jones et al., 2002).

A related post-training attitude construct is *mastery (goal) orientation* (Ames, 1990; S. L. Fisher & Ford, 1998) which is thought to be present when learners characterize successful learning in terms of making individual progress in skill development and deep understanding of cognitive content. By contrast, *performance-(goal) orientation* is associated with learners being influenced by external performance indicators (e.g. value-laden role-play feedback, grades,

etc.). While a post-training mastery orientation has been found to be related to inter-personal skill-maintenance activities and performance (C. K. Stevens & Gist, 1997), further research is needed to confirm this relationship, both generally (Alvarez et al., 2004) and, specifically, for Computer-Supported Social Learning.

Further Outcome Variables.

Compared to traditional lectures and the reading of books relevant to gaining conceptual knowledge, offering situative anchors and opportunities for application and personal reflection may also lead to higher training *satisfaction* and *motivation* in learners as well as more personalised information XXX. Both satisfaction and motivation may be treated as conditional or outcome variables. From an impact evaluation point of view (e.g. Mohr, 1995), however, it may be worthwhile to treat them as subobjectives. Possibly, prime educational goals of programmes in higher education are the support of gains in knowledge or ability or critical thinking in students. If this is the case, motivation and satisfaction can be conceptualized as merely instrumental to the prime outcomes of interest. This is also generally in line with Alvarez et al.'s (2004) model who suggest that affective reactions to training may be unrelated to training outcome and learning (cf. Alliger et al., 1997, as cit. in Alvarez et al., 2004, p. 395) while they may serve as constructs in the assessment of training content and design.

Interestingly, in their review of computer-supported medical education research, Chumley-Jones at al. (2002) noted that hardly any study addressed higher-level outcome variables such as *learning efficiency* (e.g. score gain per unit of study time) or *cost-effectiveness*.

DIGEST 2.6

Despite the general lack of research aiming at evaluating the efficacy or effectiveness of employing computer technology for social skill development, the results of exemplary studies are reported. The results suggest that the employment of computer technology may lead to variable effects on self-efficacy, and generally positive effects on subsequent performance and situational judgement. Possibly, the inclusion of computer technology in blended learning delivery formats may also result in higher learner satisfaction rates and may contribute to improved learning efficiency and/or cost effectiveness.

2.7 Evaluating Computer-Supported Social Learning

As outlined above, little developmental evaluation research exists which could help shape effective mediation training programmes beyond merely demonstrating that conflict resolution training can be worthwhile (Deutsch, 2000). A similar conclusion can be drawn in regard to research evaluating the conditions and outcomes of blending "traditional" soft skills training with contemporary technology-based hypermedia training modules. Thus, more evaluative research is warranted in both respects. Yet, approaches to the evaluation of (traditionally delivered) training on the one hand and to technology-based interaction on the other may not be easily combined, as will be outlined below.

In the following, various approaches to evaluation and their respective potentials to serve as a framework for the purpose set out above will be reviewed. It is well beyond the scope of this study to review the various theories of evaluation, however; this has been done impressively elsewhere (Shadish, Cook, & Leviton, 1995). Nevertheless, it seems important to explicate two basic tenets that underlie the following review. First, it is assumed that in the context of a higher education environment, the target of evaluation is a programme, not a product. And, second, it is believed that evaluation practice differs from the practice of research typically carried out in institutions of higher education.

Programme, not Product.

Baumgartner (1999) argues that any approach to instructional evaluation is linked to the underlying theories of evaluation and learning. For instance, the traditional textbook definition of evaluation (research) as the assessment of social intervention programs (Chelimsky, 1997; Gredler, 1996) tends to exclude other forms and fields of evaluation such as product evaluation, a point also often underscored by Scriven who, for this reason, coined the term "evaluand" (1973) to encompass a broad array of evaluation objects. In a similar fashion, Wottawa & Thierau (1998) pointed out that, in principle, anything can be evaluated. They summarized the potential objects of evaluation under nine headings which can easily be adapted for educational evaluation (see Table 10).

Table 10

Groups of Evaluation Objects

Group	Focus/Description	Examples from Educational Contexts
People	Behaviour, performance, attitudes, knowledge of individuals and/or groups (states/traits/changes)	• Educational Assessment, Certification • Knowledge Transfer after Training
Products	Properties, usability, merits, and/or comparative efficacy of educational products	• Usability of computer-based training softwares
Set Objectives / Budgets	Effects, influence, merits, impact, and/or repercussions of setting objectives	• Influence of instructional/learning objectives/targets on feedback processes
Systems/Structures	Effects, influence, merits, impact, and/or repercussions of systems and structures	• Private vs. public schooling
Environmental Aspects	Properties, influence, merits, impact, and/or repercussions of environmental variables	• Quality of learning in old and new lecture rooms
Techniques and Methods	Influence, merits, impact, comparative efficacy, and/or repercussions of techniques and employed methods	• Comparative efficacy of two methods of instruction for reading • Effects of presentation methods on retention level in the audience
Projects / Programmes	Properties, processes, merits, impact, repercussions and/or (comparative) efficacy of educational programmes and projects	• Curriculum effectiveness • Cost-effectiveness of a corporate training programme
Research/Evaluation Studies	Properties, processes, methodologies, merits, impact, repercussions and/or (comparative) value of research and evaluation studies	• Meta-analyses

Note. Based on Wottawa & Thierau (1998), pp. 60-61.

As the title of this thesis may suggest, the object of evaluation in this study can be made out to be a specific technique or method, namely Computer-Supported Social Learning (CSSL, see section 2.5, as compared to traditionally delivered social skills training). However, for reasons outlined below, it seems worthwhile to concentrate the instructional/training programme as a whole was made the preferred object of evaluation. In the following, consequently, important aspects of educational programme evaluation will be discussed with specific reference to criteria often used to determine the value of instructional programmes.

Evaluation, not Research.

Gredler (1996) points out that evaluation is often confused with educational research albeit both differ in various ways. While educational research attempts to build and tests theories generalizable across space and time, evaluation is "the process of determining the merit, worth and value of things" (Scriven 1991, p. 139) which are and remain specific. The generation of knowledge or insight may be a strong feature of evaluation, but it may also be oriented towards decision, development, or accountability (Rossi, Freeman, & Lipsey, 1999, p.

40). Moreover, controlling for construct and variable relationships under investigation, educational research employs strong research designs to lower the influence of individual preferences and values of research participants and stakeholders. By contrast, evaluation research actively and purposely "collects, clarifies and verifies" (Scriven, 1991, p. 5) stakeholder interests and values and links them to the very objects of evaluation, as the audience typically is not a scientific community or discipline, but "an identified set" (Gredler, 1996, p. 15) of decision makers or stakeholders. Payne (1994) classified the various approaches to educational evaluation according to their relationship between the evaluator and programme stakeholders:

- In so-called *Management Models*, the evaluator's role is to help inform key stakeholders (such as programme management) to make decisions about their programmes.
- In *Judicial or Adversary-Oriented Models* of evaluation, evaluators are appointed by antagonist stakeholders to limit evaluation bias.
- *Anthropological Models* of evaluation seek to identify and include programme stakeholders which are seen as contributors to a unique "culture" that should be systematically studied and addressed. And, finally,
- *Consumer Models* address potential users of a programme as primary audience. Here, programmes are viewed as "products" with a certain impact, merit and worth to be determined.

From a related perspective, Baumgartner (1999) noted that evaluations are oriented towards the interests of the stakeholders and identified two additional approaches to educational evaluation, namely,

- *Expert Models* in which specialists or authorities offer their (often subjective) professional opinion, connoisseurship or expertise on the subject under evaluation. Regularly, the relationship between these experts and the programme stakeholders remains hidden; and
- *Participant Models* or "naturalistic evaluation" which specifically address actual users of the programme as the primary audience of interest.

In contrast to many evaluation studies, social science research is often undertaken in well-controlled settings (and the title of this thesis might possibly imply that this was also the case in this project). Most mediation or social skill training activities (and possibly most blended delivery models) can be found in practice settings and thus, in highly variable and dynamic educational contexts. Here, evaluative endeavours cannot always produce information of optimal theoretical validity or accuracy; rather, their utility for generating shared meaning and decision making among stakeholders creates value. The title of this thesis relates to this

conception of evaluative research whose intended primary audience had been the decision makers and programme planners of the specific mediation training curriculum. The evaluation project's capacities for knowledge generation, for the generality of its findings, and for answering research questions of the larger community of educational researchers will thus have to be determined. As such, this study can also be thought of as a feasibility study exploring the boundaries of curriculum evaluation in a practice setting and the extent to which this type of evaluation may yield information useful for applied educational and/or psychological research.

Type of Evaluation.

Perhaps the most prominent distinction of types of evaluative activities is Scriven's (1991) distinction of *formative* versus *summative* evaluation (cf. Gustafson & Branch, 1997). In education, formative evaluations typically focus on processes or activities of the on-going educational programme or the learner. Their purpose is to help the programme planners or the learner improve their performance and/or their instructional or learning activity. By contrast, summative evaluations focus on impact and/or outcomes. Their purpose is to provide a summary judgement on the merits and "critical aspects of a program's performance" after it has ended. As far as technology-supported or blended learning is concerned, summative evaluations are purposefully applied to establish the worth of the blend in the context of its use (cf. Alexander & Hedberg, 1994). This question becomes prominent especially when technology-supported elements are introduced to extant learning systems as is currently often the case in higher education (cf. Draper, Brown, Henderson, & McAteer, 1996). This is the main reason why, in the following, approaches to training outcome evaluation are considered first.

These two types of evaluation represent ends of a continuum, not alternatives, along which evaluation processes can be distinguished (Leathwood & Phillips, 2000; Rothman, 1997b). In higher education, for example, curricular programmes are typically delivered to more than one student cohort while learning content and conditions, instructional methods and media can be subjected to gradual change. When viewed from the perspective of programme planners, results of summative evaluations of student cohorts can be used "formatively", i.e. as inducements to modify various aspects of the curriculum (cf. e.g. Powell, Hunt, & Irving, 1997; Seiffge-Krenke, 1981). Moreover, this very aspect of summative judgements contributing to goal formation may be especially fruitful for conflict resolution training programmes (Rothman, 1997a).

Problems of Goal-Based Approaches to Evaluation.

Today, various concepts and methodologies of educational project and programme evaluation exist. Cook (1997) notes that the development of general evaluation theory has come to a standstill in the early 1990ies. However, as Eseryel (2002) notes, most of the current approaches to instructional outcome evaluation have their roots in systematic approaches to instructional or training design methodologies, with *goal-based* and *systems-based* approaches being used predominantly (Phillips, 1997a).

In goal-based approaches to evaluation, areas or levels of effectiveness of training programmes are pre-defined. Possibly the most influential – and most frequently cited (Alvarez et al., 2004, p. 388) – *goal-based* approach to instructional evaluation, Kirkpatrick's (1994) *Four Levels*, was formulated and developed for business-related contexts. According to the model, evaluation should move sequentially through four levels commonly known (cf. Alliger, Tannenbaum, Bennett, Traver, & Shotland, 1997) as

1 - *Reaction* (i.e. affective reactions and utility judgements)
2 - *Learning* (i.e. changes in immediate knowledge and knowledge retention, as well as skills and attitudes),
3 - *Behaviour* (i.e. transfer to the target or application domains), and
4 - *Results* (i.e. impacts on the organisation, the larger context or the environment).

Various authors have suggested to expand this basic model, as Alvarez et al. (2004) note in a recent review. As an example, they refer to Tannenbaum et al.'s (1993) model who added 'attitudes' at post-training and divided level 3 (behaviour) into training performance and transfer performance. Recently, Phillips (1997b) suggested to add a fifth level to the model to calculate its cost-effectiveness and return of investment.)

It has also been suggested that Kirkpatrick's approach may also be applied to technology-supported learning (Bastiaens, Boon, & Martens, 2004; Henninger, 2001; Singh, 2001) and that this can model serve as an initial framework for evaluating university-based programs (Fricke, 2002). However, the point to make here is threefold, when the perspective of an impact analysis researchers (e.g. Mohr, 1995; e.g. Oliver & Harvey, 2002) is taken into account.

First, educational or training programmes are expected to solve a problem, i.e. "the program can realistically be expected to do something about it" (Mohr, 1995, p. 14). While some outcome dimensions are inherently valued, others can be conceptualized as instrumental

to these, i.e. they serve as subobjectives to the outcomes of interest. The goals predefined by Kirkpatrick's approach may therefore not adequately represent what an educational programme intends to target at. For example, a prime outcome of interest in higher education typically is building a knowledge foundation (less often skill building), and there is often little reason to expect that university-based educational programmes are either directly intended to solve societal or labour market-related problems (albeit they may or may not have such beneficial side-effects) or that pay-offs for the university itself (e.g. climate, enhancement of staff teaching skills) are primarily intended.

In addition, however, the vested interests of higher education stakeholders such as teaching staff, students, course coordinators, faculty deans, the university chancellor, funding bodies, employers, etc. (cf. Phillips, Bain, McNaught, Rice, & Tripp, 2000, Table 1.2. on p. 1.5., for examples) may deviate from prescribed evaluative goals or impact levels. They may also implicitly follow other objectives or sub-objectives and have multiple concerns (e.g. satisfying the learners, becoming certifying bodies, etc.). All these remain potentially unaccounted for in Kirkpatrick's approach. Moreover, evaluation of the higher levels can prove a challenging enterprise for a variety of reasons. For example, the number of stakeholders of human resource programmes typically is much smaller and less varied than that of university or college programmes (i.e., faculty, students, employers, community groups, parents, society, politicians, etc.). Furthermore, while the target or application domains in work-related training evaluation may be more or less identified by the trainee's current (or future) job tasks, there is an undefined and assumably much more varied array of domains programmes in higher education target at. And even if one accepts that there may be a circumscribed applied or work-related setting, evaluating transfer to such a setting can prove highly problematic.

Second, goal-based approaches to evaluation may yield information only on impact in the pre-defined areas of interest. However, they may not (or not sufficiently) address possible ways to utilize results to improve training or instruction. In other words, Kirkpatrick's approach may not yield enough information to be used as an effective strategy for formative purposes. By contrast, stakeholders in higher education may typically be expected to be interested in factors or conditions that affect the attainment of subobjectives and the inherently valued outcome and, thus, are crucial in making the programme more effective.

And third, Eseryel (2002) notes that goal-based models such as Kirkpatrick's may not make it easier for practitioners in applied educational settings to derive and implement

evaluation methods appropriate to answer their specific evaluation questions. By contrast, Oliver (2000b) observed that currently much e-learning research and development in higher education is funded in the form of self-contained projects which require practitioners to carry out their own studies who "… may have expertise in their discipline and in teaching, but it is unreasonable to assume that they will have expertise, training, and in many cases even experience of carrying out programme evaluations (Oliver & Conole, 1998)" (Oliver, 2000b, p. 23).

In summary, then, Kirkpatrick's (1994) *Four Levels* model may serve (only) as a blueprint for thinking about impact for both amending the (summative) outcome or impact goals and adding (formative) evaluation questions - based on the analysis of stakeholders and/or addressees of the evaluation.

Alternatives.

In addition - and by contrast to goal-based approaches -, *systems-based* approaches to educational evaluation originate from models of the very system to be evaluated. Its conceptualised components, relations, processes and feedback loops guide the ways to selecting and implementing appropriate evaluation methods and to the utilization of results for the improvement of instruction. One prominent - albeit somewhat dated - example is Stufflebeam's (1972) *Context-Input-Process-Product* (CIPP) model; a more recent example (Fricke, 2002) is based on Reigeluth's (1983) approach to Instructional Design who differentiates between *instructional methods* (i.e. pedagogy, media, learning environments), *instructional conditions* (i.e. variables of the learner as well as learning content), and *instructional outcome* (i.e. results and summative effects). The inclusion of conditions in particular may avoid the summative bias of goal-based approaches discussed above.

Interestingly, Fricke (2002) adds two other basic elements which provide criteria and objectives for evaluation, namely underlying *theories of learning and instruction* and *prerequisites* (such as financial resources, political guidelines, and cultural preferences) (see Figure 10 below). Relating these elements to each other results in a framework that allows for differentiating between up to nine different paradigms of instructional evaluation. For example, a given learning environment (*method*) can be evaluated in terms of its groundedness in the underlying instructional theory, an approach described by Land & Hannafin (2000). Likewise, theories of learning and instruction can be used to evaluate as to what extent personal

characteristics of the learners (e.g. learning styles as an *instructional condition*) have been accounted for.

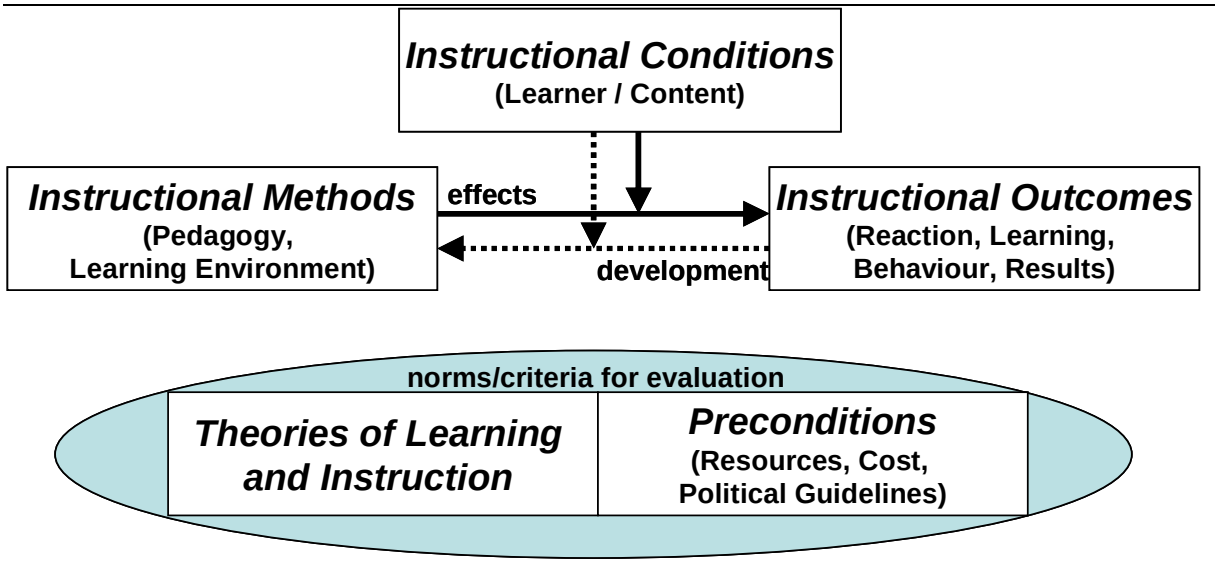


Figure 10. Paradigm for Evaluating Multi-Media Learning Environments (adapted from Fricke, 2002).

Based on the notion that an evaluation of the effects may guide development and implementation of a specific learning programme, a simplified, outcome-oriented structure is presented in Figure 10 which may be used as a frame of reference in the following.

The Role of Computer-Supported Learning.

In their report on e-learning evaluation, Franklin, Armstrong, Oliver, and Petch (2004) suggested that computer-supported learning differs from traditional classroom learning in five major ways and that, consequently, evaluation must take account of these differences.

(1) The first difference put forward by Franklin et al. pertains to the use of technology to support learning. The very resources (i.e. hardware and software) must be provided for, be made accessible and be configured correctly to fulfil its function. Thus, evaluation needs to take possible problems of use into account.

(2) Another key difference is that computer-supported learning "makes visible some of the processes that in traditional course development had been invisible" (Franklin et al., 2004, p. 1). For instance, through computerized recording, marking and tracking, learning processes become much more transparent than in traditional settings. Thus, evaluation may include possible effects and side-effects and make an augmented use of these records.

(3) Possibly, evaluation should - to a greater extent - include aspects of planning, design and development of learning materials as technology-supported learning may call for a higher degree of comprehensiveness and may often "demand more structured resources and support systems" (p. 3) than materials for traditionally delivered courses.

(4) Technology-supported learning may chiefly be used to draw on a richer set of options for objectives, styles, and situations than possible in traditional environments. Thus, evaluation models should also account for these potentials.

(5) And, finally, evaluation models may be required to pertain to new skills needed to make effective use of the technologies that support them.

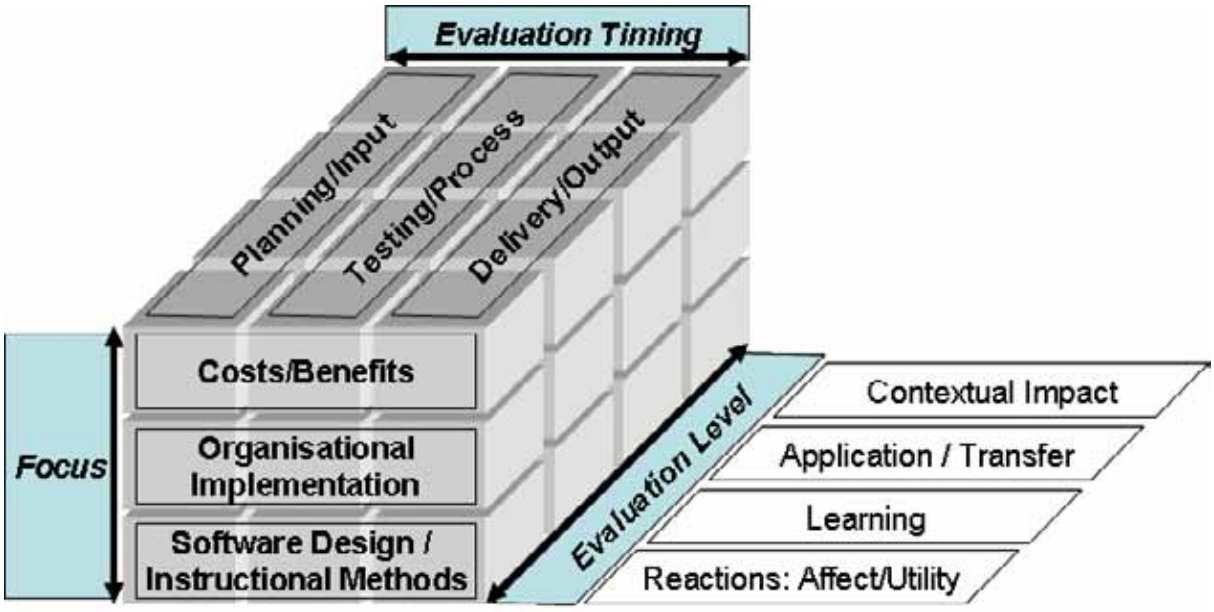


Figure 11. Evaluation Cubus (adapted from Henninger, 2001).

Similarly, Henninger (2001) maintained that, in contrast to traditional forms of delivery, computer-supported learning and instruction requires modifications of educational evaluation models. His evaluation cubus (Henninger, 2000, 2001; see Figure 11) comprises three dimensions, namely, evaluation

(1) *Timing*. i.e. evaluating planning and programme development, concurrent testing (formative evaluation), or after delivery (as in the Consumer Models discussed above);

(2) *Level* (as in Kirkpatrick's *Four Levels*); and

(3) *Focus* or the target domain the array of criteria used for evaluation attends to, i.e. methods/instructional and didactical presentation, learner and organisational variables, or costs and benefits.

Henninger (2001) suggested that, generally, the evaluation of the quality of technology-delivered training differs from the evaluation of traditionally delivered training in various aspects. Firstly, the subjective judgements of software users and thus the aspect of *learning environment usability* surfaces. This aspect is similar to Franklin et al's (2004) argument that evaluation models need to take possible problems of use into account.

Secondly, Henninger suggested that in comparison to systems of traditional delivery, "aspects of effectiveness and efficiency play a more important role in the case of multimedia

evaluation" (p. 4). This relates to the attainment of learning goals in relation to the costs of acquisition and implementation as well as time and learner effort (efficiency) and the benefits within a larger context (effectiveness). Interestingly, possibly to underscore this argument, the cost-benefit-analysis is conceptualised as a (separate or unique) evaluation *focus* rather a fifth evaluation *level* as suggested by Phillips (1997b).

However, Henninger did not provide further explanations to support his claim. While it may be true that technology-based learning requires financial resources for acquisition and/or development, it is hard to accept why efficiency and effectiveness should *generally* play a more prominent role in the evaluation of computer-based learning systems than e.g. in traditionally accepted forms of higher education or corporate training. In fact, as Mohr (1995, p. 14) points out, there is little reason to undertake or undergo training programmes unless they are expected to solve a given educational problem.

While much has been said in regard to the *Timing* and *Level* dimensions, the following discussion of the model's *Focus* dimension may contribute to conceptualizing in what ways educational evaluation should account for the incorporation of technology-supported ways of learning.

From his discussions of the evaluation model, one gains the impression that Henninger had an individual piece or product of instructional software, courseware, or a concrete computer-based learning environment in mind when he discusses possible criteria of design (such as 'authenticity' or 'quality of the examples used', p. 11). This is very much in line with the idea that (new) media or computer technologies are instructional methods and/or conditions. In fact, some researchers (e.g. Koumi, 1994; Kozma, 1994) reason that media do differentially support or mediate diverse styles of learning, its cognitive, affective, or social processes, types of learning activities, and the very instructional goals linked to performance. They essentially contend that various types of multimedia may vary in their impact on elaborative encoding, a stimulus-enriching cognitive process that eases subsequent retrieval (Najjar, 1998). Consequently, they argue in favour of research intended to examine whether learning gains can be attributed to the employment of different media or media attributes (e.g. interactivity, multimediality, video zooming and slow-motion capacities, "unwrapping" three-dimensional objects, etc.).

Others argue against such a view, despite the many research studies and meta-analysis comparing the effects and idiosyncrasies of technology-based versus traditional classroom

education. In his reviews of media research between 1912 and the early 1980s, R. E. Clark (1983, 1985) laid out his basic notion that media are mere vehicles to deliver instruction, influencing "cost or speed (efficiency) of learning" (R. E. Clark, 1994, p. 26). He suggested that media - as systems of delivery - "do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition" (R. E. Clark, 1983, p. 445). Jonassen (2001) summarised this point of view by stating: "[…] People do not learn from technologies. Rather, they learn from thinking about that which they perceive and experience. And their thoughts are mediated by activity which is mediated by attributes […]" or what he called "presentation modes" (p. 42-43).

In a related strain of arguments, Draper stated that computer-supported learning "does not cause learning like turning on a tap, any more than a book does" (Draper, 1997, p. 33). According to him, evaluation of computer-supported learning should focus on the combined effect of a product embedded in a (traditionally delivered) course environment, i.e. the entire learning package, not on one of its elements. Draper (1997) presents an interesting analogy: Attempting to evaluate an individual piece or product of technology-based learning (or one instructional method, respectively) is

> "… like considering the question 'is the 9:30 Glasgow-Edinburgh train good for getting to Edinburgh?' It is possible to imagine that there could be something uniquely good or bad about that train and not others, but in fact usually the important factors are not the details of the train itself but how it fits into people's overall travel needs and plans. People only use trains as part of wider plans, and trains are mainly good or bad to the extent that they fit (or don't) into the success of these wider plans." (Draper, 1997, CAL is only part of an ensemble section, para. 2)

This is to say that, at least in postsecondary educational settings, there typically is no such thing as sole or exclusive computer- or web- or technology-based learning. Technology-supported delivery typically is part of an ensemble of multiple learning resources and combined delivery modes. Classroom settings and computer-assisted learning components as parts of the package may reciprocally influence each other. It is therefore unavoidably and inescapably the "whole package" (including the evaluation itself) that is evaluated in terms of learning gains, cost, or efficiency. Researchers attempting to focus their summative evaluation study independently on the technology-delivered components run into seven potential problems discussed by Draper (1997). His arguments are reformulated in the following:

(1) Unless made compulsory, technology-delivered programme components may not be used at all by students. This refers to the critical role of motivation for learning gains which may be influenced by classroom settings.

(2) Typically, aspects of the technology-delivered programme components are discussed and referred to in face-to-face contact with teachers and peer learners and thus need to be included in valid evaluations.

(3) This refers not just to the dialogue based around learning contents, but also the attitudes to the employed technology itself voiced in classroom settings may strongly affect learners' use of technology-delivered programme components (HALO-effect).

(4) The study strategies of learners are known to influence outcome. While these can be heavily influenced by face-to-face contact in the classroom, technology-delivered programme components presently do not elicit stable study strategies in students.

(5) When focussing solely on technology-delivered programme components, all outcome evaluation is prone to unnoticeably include what Campbell and Stanley (1966) called "history effects", i.e. the learner's gains depend on alternative learning resources and student activities.

(6) In fact, just as poor teaching in universities may be masked by compensatory strategies by students, learning gains may be affected little even when ineffective or useless technology-delivered training is employed.

(7) Finally, reactive effects of the evaluation itself (often embedded in technology) may affect the results produced, e.g. through Hawthorne effects or selection bias (e.g. participants feel valued, attend significantly more to the technology-delivered programme components and its contents) or testing effects (e.g. tests affect learner's attitudes, priming effects due to pre-tests, etc.).

In the light of these arguments, a valid impact or outcome evaluation cannot be implemented when individual pieces of technology-based learning are focused on. (For an in-depth discussion of educational software and courseware evaluation see Baumgartner, 2002.) Possibly, Henninger had the combined effect of the "course package" in mind if package is understood to be the "learning environment" – or, alternatively, what has been referred to as "organisational implementation" in Figure 11, a second focus if this is understood as focusing on the "mix" of programmes and delivery modes offered by the organization.

"Integrative Evaluation".

In summary, Draper (1997) suggests that it is virtually impossible to (summatively) evaluate the impact of computer-assisted learning in higher education as any outcome parameter inescapably reflects to a blend and interplay between the technology-delivered and traditionally delivered components and their evaluation. A given mix of delivery modes can, as a total or "final" product, duly be focused on during evaluation as suggested, for example, by Bastiaens, Boon, & Martens (2001); however, doing so would mean that the instructional methods are treated as one single independent variable (mediated by instructional conditions as in Figure 10). Whereas this approach does not preclude developmental or formative aspects that help change the instructional method or educational programme (as Figure 10 correctly

implies), due consideration indicates that evaluators may want to obtain ideas on how to improve integration of delivery modes or modifications of the mix. This may be true especially when technology-supported materials are being introduced to extant courses or programmes, as Draper, Brown, Henderson, and McAteer (1996) suggest.

These ideas mirror the core lines of thought put forward in the so-called Integrative Evaluation model (Draper, 1997; Draper et al., 1996). It can be conceptualized as a management model of evaluation as it seeks to inform those responsible for including computer-supported modes of delivery. As such, it is can be understood as being focused on implementation (as in Henninger's Evaluation Cubus). However, rather than exclusively measuring various levels of impact (or outcome), Integrative Evaluation as an empirical approach aims at "providing better information than is ordinarily available" (Draper et al., 1996, p. 17) of both the effect or impact of "blended delivery packages" and, on the other hand, the methods, conditions and processes of learning (such as the context of the course, policy, resources, overt and underlying learning objectives) as these may be factors with a significant impact on learning. As such, as Oliver (1999) put it, the Integrative Evaluation approach "aims to evaluate the course's use of educational technology, not the educational technology alone" (p.13).

In his studies, Draper and his colleagues used a variety of methods, including video observations, within-group performance examinations, task experience, learning resource, and computer experience surveys, knowledge quizzes, interviews with focus groups, and confidence logs. Their approach thus incorporates a mix of evaluation methodologies used in psychology and ethnography (cf. Parlett & Dearden, 1977). The inclusion of the latter underscore the researchers' notion that their evaluation strategy is essentially situation-specific; thus, there is no primary need to generalize the findings beyond the specific population studied (Oliver, 2000a) . Vice-versa, concerns have been voiced about the use of highly controlled experimental research designs as they may lack generalizability to real settings (Draper, 1997; Draper et al., 1996).

Oliver (2000a) cautions that the methods favoured by integrative evaluation may challenge confidence in the validity and completeness of the findings when studies are carried out in the "distributed, fragmented context of distance education, coupled with the reduced reliability arising from self-selecting responses and harder sampling choices for interviews" (p. 90). This problem may be tackled in traditional higher education where face-to-face

interaction still dominates. Here, blended assessment strategies and multiple data sources (such as records of classroom interactions, performance examinations, portfolio assessments, etc.) may be used. Also, entire cohorts rather than subsamples can be included.

Another challenge for integrative evaluation has been formulated by Taylor, Woodman, Sumner, and Tosunoglu Blake (2000) who suggested that, theoretically, an integrative evaluation strategy has to "encompass both the content of the course and the ability of students to understand the teaching, as well as the way that media had been used to communicate those teaching aims" (p. 46). Questions, for example, should be posed in a manner that encourages learners to focus their response on the deeper role of the various media delivery modes for educational effectiveness rather than leading them into adopting "a quasi-analytic 'media critic' role that focuses on superficial elements of a medium", such as "aesthetic aspects of the interface in multimedia systems" (ibid.). Taylor et al. therefore advocate developmental testing and the refinement of questionnaires.

Gunn (1996) argues that Draper et al. do, in fact, maintain a connection to traditional scientific methodology and that "…believers in scientific experimental methods consistently refuse to believe that there may be another, more appropriate way" (p. 158). She recommended avoiding the goal of producing generalizable results altogether and suggested an alternative, situated evaluation approach with a narrow focus and a loose (?) combination of methods based on action research, case study methodology, and grounded theory development. This approach may be easier to follow when formative objectives are prevalent (or followed exclusively). However, this approach seems like throwing the baby out with the bath-water when viewed from a summative perspective. In higher education, evaluation may be viewed as an on-going or repeated undertaking, and the circumstances allow for, at least, some hope that generalization to the target populations of current and future students may not be so far-fetched. In sum, therefore, Draper et al.'s Integrative Evaluation model seems to be a viable compromise when formative aspects are not the sole focus of evaluation.

Summary.

When the focus is on outcomes, outcome criteria may be considered in term of explicit and implicit learning theories. The social learning framework laid out above may therefore justly serve as a theoretical base for the development of evaluation criteria. In addition, both instructional methods and conditions should be accounted for in the evaluative design.

Moreover, the choice of evaluation design may also be mediated by the inclusion of learning technology when two main arguments are considered. Firstly, it seems reasonable not to attempt to evaluate simply the effects of an educational technology as it is almost always embedded into traditionally delivered components. Thus, the evaluation should focus on the social learning programme as a whole, i.e. on course and/or curriculum level, not on technology alone. And, secondly, as typically technology-mediated elements are currently added to extant courses or programmes in higher education, formative aspects may inescapably play an important part for programme management and teaching staff. Both arguments are reflected upon in the Integrative Evaluation approach outlined above.

DIGEST 2.7

Evaluative research is warranted on the impact of mixed delivery skills training. It is explicated that various differences exist between research and evaluation, and that the objectives of this study are essentially of evaluative nature, with a specific focus on programme impact in higher learning. It is argued that, for two main reasons, approaches to training outcome evaluation should seek ways to incorporate aspects of formative evaluation. On the one hand, computer-supported learning differs from traditionally delivered training in major ways. For example, extended evaluation systems may take account of problems and effects of computer use and tracking processes, of preconditions, of potentials unique to technology-supplied materials, and of additional skills required. And, on the other hand, outcome evaluation of newly introduced blended learning scenarios typically stimulate refinement of the blend to be introduced by programme management and/or teaching staff. It is suggested that Integrative Evaluation (Draper and his colleagues, 1996, 1997) is a viable approach for blended learning scenarios.

2.8. Conclusion and Research Questions

Summarizing notions and findings of the research outlined above, it is a worthwhile effort to shed some more light on the question if (and, possibly, how and under what conditions) technology-enhanced socio-communicative competence development "works".

Present-day computer technology allows for the development of software in general accordance with Social Cognitive Theory and related theories of learning. Respective software in line with the framework model of socio-communicative competence development outlined above would provide situated, problem-based examples for socio-communicative interaction with open-ended questions, multiple-choice judgement exercises, modelling, feedback, and written reflection, similar to Gentry's (1992) example.

Such software cannot replace direct, situated interaction traditionally offered in classroom-based trainings; it may, however, as tailored courseware be incorporated into or (or 'blended' with) such trainings, in the same manner as preparatory or supplementary written materials or textbooks are traditionally used to impart or strengthen conceptual knowledge. This approach has been termed Computer-Supported Social Learning (CSSL, see section 2.5).

Theory suggests that the provision of courseware based on CSSL principles could effect in changes in an individual's cognitive-affective schemas used for both the judgement of a professional type of situation, the corresponding internal cognitive-affective appraisal as well as the generation and execution of behavioural responses. Concurrently, conceptual knowledge as typically imparted in books or lecturettes may be made accessible and consolidated and enriched at the same time by providing vivid, situated 'anchors'. Consequently, this notion would advocate the use of CSSL courseware as an additional preparatory means (i.e. 'between' textbooks and training) or as a means of post-training development (or skill maintenance).

Thus, the impact of CSSL may be best evaluated in 'blended' learning environments. This is in generally in line with current approaches to the evaluation of technology-supported training. For example, Draper (1997) suggested that outcome parameters of summative evaluation inescapably reflect the blend and interplay between the technology-delivered and traditionally delivered components and their evaluation. As far as the provision of the computer-based software is embedded within a (traditional) university-based course or programme, a viable evaluation strategy following the "Integrative Evaluation" model will seek to identify specific outcome and impact evaluation parameters for "course packages" and the

entire "programme" rather than attempting to evaluate the new technology-supported element alone.

The principal question may thus be put as follows: *Does the introduction and learner's use of the CSSL courseware, combined with the pre-existing preparational materials and soft skills training, result in augmented increases and higher resulting levels of socio-communicative competence?*

The framework model of socio-communicative competence development also suggests that, whereas conceptual knowledge may be assessed by knowledge tests, situated knowledge, skills, attitudes, and other attributes (KSAOs) of socio-communicative competence may only be tapped indirectly, i.e. by use of performance assessment, situational judgement testing, or an analysis of situative social-cognitive processing or attitudinal self-reports (e.g. self-efficacy).

In general, exactly these are the methods and outcome parameters used in evaluative studies. Possibly, a direct assessment of situated behaviour through performance assessment may be the 'yardstick' or benchmark against which the results of situational judgement or cognitive-emotional processes and attitudes are to be referenced. However, it is also true that use of this method is expensive and probably currently not feasible in higher learning where, typically, a great number of candidates are to be assessed. Therefore, for the time being, the performance assessment approach to socio-communicative competence was shelved and excluded from the eight assumptions or 'hypotheses' presented in the following.

Centring around the principal question outlined above, these assumptions were based on theoretical considerations and empirical findings detailed previously as well as intense discussions with programme management and stakeholders. Among other aspects predominantly focusing on exploratory issues and formative aspects (such as curriculum development, see below), they were also used as guidelines for the evaluation of impact and subsequent data analyses.

H1 The introduction and learner's use of the CSSL courseware, combined with the pre-existing soft skills training, is expected to result in augmented increases and higher resulting levels of participants' situational judgement skills.

H2 The CSSL courseware will supplement and enrich extant preparational materials (i.e. textbooks and recommended texts) in such a way that complementary use of the courseware results in superior situational judgement skills.

H3 The introduction and learner's use of the CSSL software is expected to result in higher conceptual and taxonomic knowledge levels.

H4 The introduction and learner's use of the CSSL software is expected to result a higher level of interest in the subject of mediation and conflict resolution at the end of the curriculum.

H5 The introduction and learner's use of the CSSL software is expected to result in higher learning gains in areas addressed.

H6 The introduction and learner's use of the CSSL software is expected to result in higher increases and higher resulting levels of participants' self-rated confidence in theoretical understanding and knowledge as well as in their confidence to apply socio-communicative skills used in mediation.

H7 The CSSL courseware will supplement and amend extant preparational materials (i.e. textbooks and recommended texts) in such a way that complementary use of the courseware results in superior increases and higher resulting levels of participants' self-rated confidence in theoretical understanding and knowledge as well as in their confidence to apply socio-communicative skills used in mediation.

H8 Increased duration, intensity and thoroughness of CSSL are also expected to result in a more accurate level of the participants' self-rated confidence in their theoretical understanding and knowledge of applied socio-communicative skills for mediation.

It should again be pointed out that these were not the only questions and assumptions guiding evaluation. In the course of this action-oriented project, various other questions were explored and preliminary results concerning both summative and formative aspects were feed-backed to programme management and peer-tutors to support learning and change management. The following, therefore, should be understood to be one (major) part of a greater whole.

DIGEST 2.8

Theory and research suggests that the outcome of planned socio-communicative competence development traditionally based on face-to-face interaction may be accompanied by advanced learning technology. In the same manner as written materials and/or lecturettes are frequently used to impart or strengthen conceptual knowledge thought to affect an individual's socio-communicative KSAOs, software based on principles of Social Learning may make these conceptual models more accessible, providing opportunities for situated and situated, problem-based experiential learning.

Yet, what impact on socio-communicative competence development can be expected if such software is incorporated into (or 'blended' with) hands-on training as preparatory, tailored courseware – an approach that has been termed Computer-Supported Social Learning (CSSL, see section 2.5)? Eight assumptions were formulated to be used as guidelines for the evaluation of impact and subsequent data analyses.

3. Rationale and Method

In the foregoing analyses, mediation training was viewed from the perspective of socio-communicative competence development. In line with this view, approaches to conceptualizing socio-communicative competence, its assessment and development have been reviewed, explicated, and placed into the context of current educational research as well as more general perspectives on learning.

An integrative model for understanding socio-communicative competence development and integrating educational technology into the developmental processes of socio-communicative competence has been put forward to guide research, practice, and evaluation. Finally, after a review of evaluation models for Computer-Supported Social Learning (CSSL) and its outcomes, evaluation questions for this study were presented. These essentially centre around one issue: Does the introduction and learner's use of the CSSL courseware, combined with the pre-existing preparational materials and soft skills training, result in augmented increases and higher resulting levels of socio-communicative competence?

In this chapter, the study's rationale and method will be presented. First, a detailed account of the setting and background of this study is offered, followed by comprehensive descriptions of the evaluation design, the study participants, instruments used as well as procedures undertaken to generate information of such a kind that support getting to the bottom of the evaluation questions outlined above.

3.1 Setting

The "Seminarreihe zur Problem- und Konfliktberatung" (Counselling for Facilitated Problem Solving and Mediation and Conflict Resolution Training Programme; "*PROCON*" hereafter) is a curricular series of training modules offered to graduate psychology students at the University of Hamburg's Department of Psychology.

Since the mid-1990s, *PROCON* has been established, run, and administered by the *Arbeitsgruppe "Beratung und Training"* ('Consulting and Training' Work Group; "*BuT*" hereafter) within the Department's Educational Psychology division. *PROCON* is a voluntary add-on for graduate students interested in enhancing both their theoretical knowledge about conflict resolution as well as their practical skills in working with groups and teams in which problems and disagreement prevail. In correspondence with the findings of earlier studies (e.g. Deppen, 2001; Kilburg, 2005b), students accepted for participation in the curriculum ideally have at least basic knowledge and experience in group facilitation techniques (e.g. agenda setting, visualization, use of sort cards and forced choice stickers, brainstorming; cf. e.g. Doyle & Straus, 1976; Nitor, 2002) as well as basic one-to-one-counselling and communication skills (e.g. active listening, confronting, psychodramatic doubling, cf. e.g. Jacobs & Redlich, 1998; Schulz von Thun & Bossemeyer, 1993).

Interest in voluntary participation in the *PROCON* programme has proliferated in the beginning of this decade. Annually, up to 100 sincere inquiries concerning application and participation are received from students of psychology and other fields of study.

The *PROCON* Curriculum.

Originally consisting of only one peer-tutored module in which students explored strategies to resolve conflict in role-played organizational teams (Redlich & Elling, 2000), the need and demand to introduce further modules was soon realized by students and instructors alike. By 2001, the *PROCON* curriculum had a duration of two semesters (or one year) and consisted of two modules concentrating on enhancing knowledge and skills needed for facilitating problem solving in groups, two modules promoting basic knowledge and skills required to act as a mediating party in both two-person and small group conflict scenarios, and one final module to help students and experts compare different approaches to problem and conflict resolution.

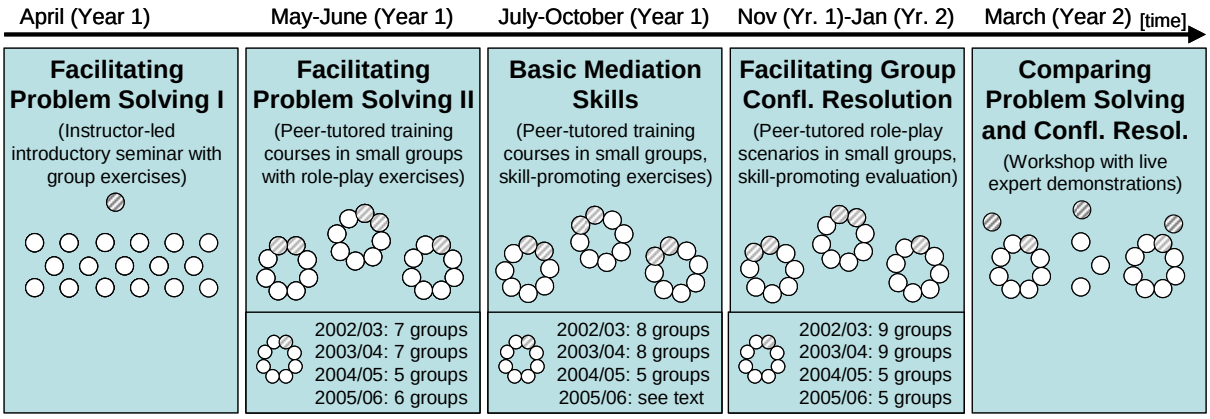
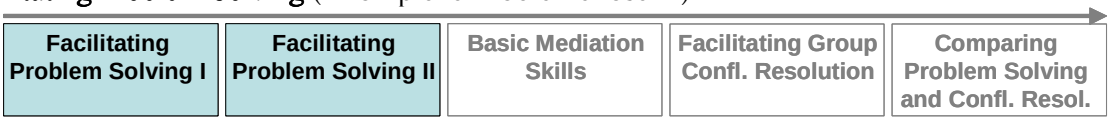


Figure 12. ***PROCON*** **Core Curriculum.**

The *PROCON* curriculum has been subject to continual changes due to budget restrictions as well as conclusions drawn from evaluative feedback from both students and advanced students serving as peer tutors in the training courses. During the course of this evaluation study, for example, additional modules were added at the beginning and end of the existing *PROCON* curriculum, namely, a module aimed at developing participants' facilitation skills and a module on multi-party mediation and between-group conflict scenarios. In the last cohort included in this study, time and budget restraints led to additional changes in various modules (as described below; see also Figure 13 in section 3.3 and Figure 16 in section 3.4 for a detailed overview). Also, module preparation materials and the pre-devised simulation scenarios have been constantly updated. However, the basic or core curriculum – as summarized in Figure 12 – has essentially remained unchanged for several years.

In the following, the instructional system and the main topics covered in the curricular modules are outlined before the more general pedagogies and the instructional approaches will be detailed.

Facilitating Problem Solving ("Komplexe Probleme lösen")



The problem management strategy approach taught in the first modules focuses at facilitating the understanding and resolution of structural, factual, physical or material problems. As Kilburg (2005b, p. 33) notes, the approach targets in particular at complex problems the (non-)resolution of which affects numerous other areas and the ramifications of which are difficult to predict.

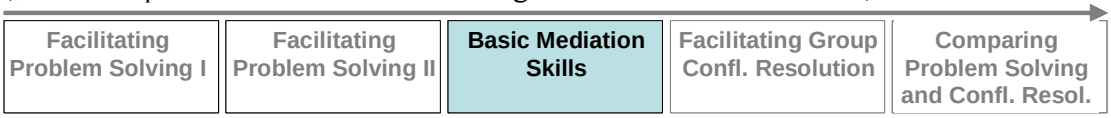
Typical case-study scenarios, for example, include shortcomings in filament production processes in a chemical corporation, sales problems of a car manufacturer due to repeated car lamp malfunctions, or a table tennis bat producer's troubles with different rubber materials and glues used for the blades. Exercises like these are used to engage

students in applying the strategic approach from the perspective of a consultant in both one-two-one and one-to-group settings.

Generally based on an approach described by Spitzer and Evans (1997) and developed further at the University of Hamburg (Burkhart, 1982; Ueckert, Knop, & Burkhart, 1982), the two-day introductory seminar focused on acquainting the students with stages of the strategy (i.e. *Problem Analysis*, *Root Cause Analysis*, *Solution Scenario Generation*, *Decision Analysis*, and *Potential Problem/Opportunity Analysis*), whereas the subsequent three-day peer-tutored training courses in small groups (8-12 students) concentrated on fostering the efficacy of students in strategy application.

Basic Mediation Skills

(„Kernkompetenzen der Konfliktberatung/Zwei-Personen-Konflikte“)



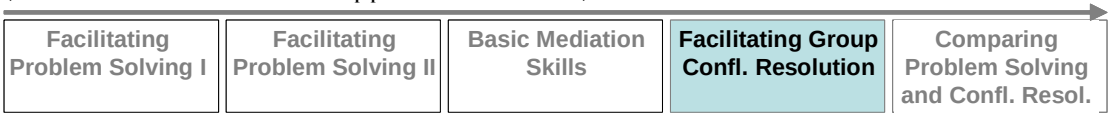
In an individual preparation phase preceding this module, curriculum students acquainted themselves with concepts intended to raise their awareness and perceptiveness of basic communicative interventions of mediators as detailed by Redlich (2004a; see also below). These include (a) the *structuring* of the mediation process, (b) deepening understanding through interviewing and two-sided *active listening*, (c) the *joint development of solutions* through brainstorming, and (d) *depolarizing*, *i.e.* the promotion of fairness through incremental opposition to fouls of language use.

During the three-day training courses, peer-tutors led small groups of approximately 10 students through a series of lecturettes dealing with the use of these basic interventions and issues complicating their application throughout the mediation process as described by the strategy model. These issues were experienced and further explored through repeated practical exercises and role-play simulations (both pre-devised and authentic on-the-spot contributions were used).

In the last cohort included in this study (2005/06), this module was converted to an instructor-led, three-day training course which focused on mediation or couple counselling for dyadic conflict in personal relationships. The basic rationale for this training course is detailed in Thomann (1998) and Thomann & Schulz von Thun(1988).

Facilitating Conflict Resolution in Groups

(“Interessenkonflikte in Gruppen aushandeln”)



Redlich & Elling’s (2000) training manual containing both case simulation scenarios as well as techniques for role-play, feedback, reflection, and evaluation served as a blueprint for this module. Most of the scenarios published had been based on authentic cases (i.e. modelling real world, cf. Shaffer & Resnick, 1999) adapted from interviews with trainers and consultants involved in human resource and team development measures. For the purposes of training, all cases had been adapted for training purposes. As such, the pre-devised scenarios focused on a group or team with six to eight members in an organisational setting and had a reasonable level of complexity.

During first, half-day group course meetings, peer-tutors assigned four to five pre-devised mediation cases to a same number of pairs (or 'tandems') of students taking part in this module. The assignments were set through role-played contracting and involved realistic in-group conflict scenarios.

All cases were intended to pertain to different strategic aspects and stages of the mediation process. Typically, one of the four cases focused on the first in-group facilitation phases as described by Redlich (2004a, 2006), namely *Contributing to Contact* and *Identifying Conflict Issues and Reaching an Agreement on Conflict Resolution Procedures*. *Clarifying perspectives and underlying interests* was the target phase of one or two further cases, and in one to two scenarios, the main mediation task was *Cultivating positional shift and finding agreements* and/or *Safeguarding transfer and implementation*. Details on the latter phase can be found in Rogmann and Redlich (submitted for publication).

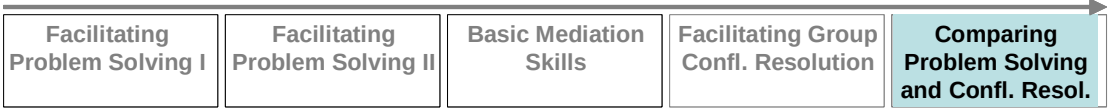
After the first meeting, the student tandems had an average of 3-4 weeks to prepare for the subsequent three-day group course meeting. During this meeting, the case scenarios were simulated, each with the prepared tandem performing as mediating facilitators and the respective remainder of the group serving as role-players. For purposes of later individual reflection, all role-plays were videotaped and converted to digital multimedia compact disks which were subsequently handed out to the respective mediating tandem.

After each role-play simulation, the peer tutors induced a careful analysis of the mediation processes. Typically, this process consisted of four consecutive phases, namely, (a) role players giving feed-back on demand; (b) the collection of a subset of 'good practice' or exemplary behaviours as presented by the mediating students; (c) brainstorming and collection of a number of difficult or crucial situations that surfaced during the simulation and (d) subsequent re-enactment of these with alternatives to the previously performed mediator behaviour spontaneously invented by students and tutors (referred to as *act-storming* by Redlich & Elling, 2000).

In 2004 and 2005, this module was complemented by a three day simulation workshop with all cohort participants, focusing on between-group conflict and intervention and facilitation techniques for large groups (cf. Stagge & Redlich, 2007).

Comparing the Consultative Practice of Problem Solving and Conflict Resolution

("Praxis der Problem- und Konfliktberatung")



In a final three-day workshop module, all curriculum students met two to three expert consultants with significant practice experience but distinct consultative approaches to problem solving and conflict resolution in organisations.

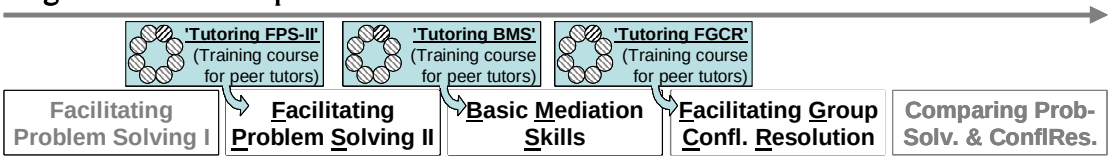
During the first day, all students were trained by all experts who, repeatedly 'circuit trained' a portion of the student group in round-robin fashion, thus, reviving the students' skills and understanding for their respective approach.

During the two consecutive days, the students were instructed to attend a number of simultaneous live demonstrations during which the experts took on the role of a consultative third party in identical same pre-devised written case simulation scenarios. As observers or partictipant role-players, the students were asked to record and

document the consultative approach of the respective expert for subsequent presentation in a plenary group setting. Questions pertaining to difficult or crucial situations that had surfaced during the simulations were spontaneously re-enacted with the various experts *act-storming* for improved contrast and comparative evaluation of their respective approaches.

Due to budget restraints, this course could not be offered to student participants of the 2005/06 cohort.

Training Course Workshops for Student Tutors



Students that had completed all *PROCON* modules were permitted to enter a second training phase, namely "train-the-trainer" (i.e. 'train-the-peer-tutor') workshops. These consist of three-to four-day workshops being led by expert instructors. In terms of the general instructional approach, the methods and didactics used, these workshops generally were replicas of the training courses for all *PROCON* participants described above, i.e., tutors assumed the perspective of their later participants in an exemplary course conducted in such a way as those courses the advanced students were being trained to tutor. In addition, the students were given access to all materials, exercises, case scenarios, were trained to lead and structure sessions, to employ methods of reflection, and to react to predicted problems.

Thus, the workshops also typically included reference and advice to prepare the tutors-to-be for their roles as coordinating and facilitating agents. Among the various predicted problems regularly surfacing in the workshops were, for example, "How to deal with latecomers", "How to contribute to an error-friendly learning climate", "How to initiate a productive act-storming", "How to give resource-based and clear feed-back". These issues were explored using experiential and reflective methods in line with (and making explicit reference to) the general instructional approach detailed in the following.

These workshops were regularly 'wrapped-up' by follow-up sessions after the conclusion of all peer-tutored trainings. Here, in general accordance with an action evaluation strategy, both the experiences of tutors as well as results obtained in module online surveys were discussed. Starting points and levers for formative changes were identified and the peer tutors were invited to share their impressions as to the 'social' validity of the evaluative findings and their possible interpretation.

DIGEST 3.1

The *Counselling for Facilitated Problem Solving and Mediation and Conflict Resolution Training Programme* (*PROCON*) is a curricular series of training modules offered as a voluntary add-on to graduate students at the University of Hamburg by the Department of Psychology's Arbeitsgruppe *Beratung und Training* (*BuT*). The curriculum has a duration of one-year an consists of five core modules (*Problem Solving I* and *II*, *Mediation Skills*, *Conflict Resolution in* Groups and a final workshop module) most of which are peer-tutored training courses.

3.2 Instructional Approach and Pedagogy

BuT - Philosophy.

Traditionally, in scientific higher learning the examination, consolidation, propagation and acquisition of knowledge and its foundations is emphasized. By contrast, the Arbeitsgruppe "Beratung und Training" (*BuT*) overseeing and coordinating the *PROCON* curriculum pursues the idea of combining science and practice. Here, much in contrast to traditional higher education, knowledge acquisition and skill building are viewed as co-equal priorities. This is in line with one of *BuT*'s other primary goals in higher learning, namely, to prepare students for professional practice through the development of key psychological skills. In congruence with this idea, the emerging instructional philosophy includes three related elements, namely, (a) the personal acquisition and re-construction of imparted knowledge, (b) the provision of opportunities for application and practice, and (c) the tying-in of knowledge acquisition and performance experiences with personal qualities of the learner (Schulz von Thun, 1981, 1998, 2004a, 2004b).

Degree of Autonomy in Participation and Learning.

As *PROCON* has been established as an 'add-on' to credited courses of study eventually leading to academic credentialing (i.e. university degrees), participation has been made voluntary. Learners intending to participate have to apply in writing, stating their motivation and qualification and experience with regard to group facilitation, counselling, and communication skills. In the cohorts covered by this study, all applicants were accepted for participation. Whilst encouraged to complete the curriculum within the one year projected, students were explicitly permitted to opt for a complete two-semester intermission or even to completely discontinue their involvement. Interest in the subject matter was largely taken for granted, and utilization of the preparation materials offered was essentially placed within the personal responsibilities of the students. Attendance of more than 85% and full participation in the evaluation survey was required to successfully complete a curriculum module and receive eligibility for enrolment in the subsequent module. In the experiential training modules led by advanced peers, the student groups themselves control the amount and emphasis placed on reflection and feed-back and, individually, the degree of personal involvement in exercises and demonstrations undertaken.

Enactive Learning.

Bandura's social learning theory (1969, 1971, 1977, 1986, 1999, 2001) can be regarded as a theoretical cornerstone of the pedagogical approach. In congruence with what Bandura (1986) referred to as *enactive learning*, principal learning ingredients here are thought to lie (a) in the behavioural action and the environmental responses and (b) in the cognitive processing of the behavioural responses and environmental feedback information by a learner.

Both elements are manifested in the pedagogical methods used in the curriculum modules. The vast majority of exercises, case scenarios, and simulations used in the *PROCON* curriculum target at third-party functions. They are devised with the very idea that students take on the roles of counsellors or consultants and actively try to put into practice adapted behavioural strategies and experience the responses of the other role players. A good example here is the "Facilitating Conflict Resolution in Groups" module, but, likewise, the other in-class training module entail an extensive amount of direct experiential, behavioural exercises aimed at an active transfer of strategies and cognitive knowledge into situated behaviour. However, enactive learning is also effectuated by peers taking on the roles of persons *receiving* counselling or consultation in that they can experience the effects of actions and interventions by the consultants.

Moreover, both module instructors as well as the trained peer students leading the training groups actively engage participants in these exercises as well as the subsequent cognitively oriented reflection in which the individual experiences and resulting questions are shared and collaboratively are classified and responded to (e.g. through role player feed-back or the group-based collection of 'crucial' situations) Cognitive processing is also promoted by individual reflection (e.g. through a personal analysis of the role-play video-tape). These individual and collaborative methods used in the modules are thought to promote abstraction, reflection, and rule generation. This implies, for example, that students are enabled to furnish with meaning the behavioural models and strategies imparted as well as the conceptual "language" (terminology or *Sprachspiel*) drawn on and referred to in preparatory materials or in-class lecturettes.

Modelling and Observational Learning.

However, a group of student role players reacting to their peer facilitator in the "Problem Solving" training module, for example, do not only serve as responding and feed-backing agents. They also try to match the behaviours displayed of the counselling party with

their own conceptions of counselling and intervention (behaviour). Thus, students do not only learn *enactively*, but also through processes of *observation* of others (students, peer tutors or experts) performing.

Three strains of influence are at the heart of additional methods applied to promote modelling and observational learning in the *PROCON* curriculum, namely, (a) altering inhibitions over previously learned behaviour, (b) facilitating responses in that they serve as social prompts creating motivational inducements for observers to behave accordingly and (c) creating behaviour that had a zero probability of occurrence prior to an observation, even under conditions involving high motivation (Bandura, 1969). For example, the above-mentioned module "Facilitating Conflict Resolution in Groups" incorporates a group-based reflection to collect 'good practice' models, i.e. samples of commendable behaviours that were displayed by the 'facilitating' tandem. The verbal explanations, discussions, and behavioural performances accompanying this collection can be viewed as influencing processes of attention (e.g. the discriminability of modelled activities, or their perceived functional value), retention (e.g. conciseness and opportunities for rehearsal), and motivation (e.g. in that this method may raise self-efficacy of the performing students). Yet, these are viewed as decisive processes for observational learning to occur (see above; cf. Bandura, 1986, 1999).

Resource-based Learning and Error-Friendliness.

A method used widely used in the curricular modules combines enactive with observational learning: *act-storming* (cf. Redlich & Elling, 2000, see above). Act-storming is employed, for example, to create and identify behavioural models for 'crucial' or 'tricky' situations through spontaneous enaction. Students, peer tutors and experts engaging in act-storming draw on resources available to them (e.g. knowledge, competencies, skills, and other attributes) to generate enactive learning experiences.

As such, act-storming in particular is a good example for the overall resource-based pedagogical "culture" (cf. Campbell, Flageolle, Griffith, & Wojcik, 2002) fostered in the *PROCON* curriculum. Students are viewed as active learners in that they are encouraged to use materials beyond those recommended. A significant proportion of the students begin their studies after having accumulated various years of professional experience in such fields as nursery and physiotherapy, business, teaching, and they are welcome to actively integrate these experiences. Moreover, both teachers as well as the advanced students employed as peer tutors

see themselves as facilitators rather than expert instructors. In the workshop training courses for peer tutors, they are helped to promote error-friendliness and an atmosphere of collaborate problem-solving through joint experimentation in their training courses.

Mutually, students, peer tutors and instructors serve as behavioural models for the observing participants. A basic tenet of the overall pedagogical approach, and of act-storming in particular, is the idea that there is not a single 'optimal' behaviour or 'right' reaction to meet the complex demands of a given social and systemic situation, but a diverse and variable array of opportune, suitable behavioural responses (cf. Schulz von Thun, 1989; cf. Schulz von Thun, 2004b). Through enactive and observational learning, learners can discover various elements of this array through trial and error and are encouraged to do so. Overall, this approach can be characterized as intended to contribute to a *mastery* rather than a *performance orientation* (Ames, 1990; Fisher & Ford, 1998; Steven & Gist, 1997) in students.

Coherence (*Stimmigkeit*) and Personal Development.

However, apt socio-communicative performance entails "dual coherence" in that it does correspond with the situational demands, but is also authentic in that it is consistent with the identity, nature, or character of the person communicating (Schulz von Thun, 1998; 2004b).. In terms of *PROCON* curricular pedagogy, the latter calls for (a) sensitive feed-back to the learner, e.g. in terms of perceived ambiguities or sensed lacks of impartiality or authenticity (b) opportunities for reflection on whether or how the consultative strategies imparted can be adapted to the individual and (c) a delineation of areas of personal development likely to contribute to a beneficial consultative performance in problem solving and mediation.

In some of the modules and in some of the training courses, peer tutors have experimented with offering a forum for voluntary solicited feed-personal back. Near the end of the training course, individuals would formulate a specific query pertaining to personal development and then self-select a group of peers to receive resource-oriented feed-back with reference to their individual question.-

A more systematic opportunity for self-reflection near the end of the curriculum is embodied in the written analysis of the student's own video-taped performance. The participants were expected to complete three reflective steps. First, in line with the notion that a focus on enactive attainments potentially raise self-efficacy (Bandura, Adams, & Beyer, 1977), they were asked to describe behavioural sequences performed by themselves which

could count as exemplary models. Second, they were expected to review their performance and portray a skill or an approach to *Conflict Moderation* (Redlich, 2004a; Rogmann & Redlich, submitted for publication; Redlich, 2003) which they regard as their next step in personal development. This assignment relates to the idea that learners potentially benefit from getting full access to observing their own behaviour as they can compare their performance against an internal conceptual standard which serves as a reference (cf. Bandura, 1986). And third, they were asked to conceptualize the latter in terms of the *Developmental Quadrangle* (*Entwicklungsquadrat*) diagram (Schulz von Thun, 1989, 2004b) to promote goal-setting (Locke & Latham, 1990; Bandura, 1988) and goal orientation (Dweck & Leggett, 1988; Ames & Archer, 1987).

'Instruction' versus 'Learning' in the *PROCON* Curriculum.

For participants, there is a comparatively high degree of freedom and autonomy in participation and learning. The curriculum managements subscribes to a view of learners as agents, and to its underlying theories of learning, and has reduced lectures to a minimum in favour of discovery-based and reflective pedagogical approaches to learning in peer-led training courses. In terms of the 3C model outlined above (see section 2.5), classroom interaction mainly focused on fostering communication and both individual and collaborative knowledge construction. In face-to-face situations, instruction in the sense of presenting or imparting facts or knowledge sets was essentially limited to a minimum. As far as 'content' is concerned, the didactical approach taken here was to ask participants to prepare study preparatory or accompanying written materials.

As far as motivation is concerned, the *PROCON* programme management can be understood to conceptually subscribe to the basic ideas of self-determination theory (Deci & Ryan, 2000; Ryan & Deci, 2000b). On the one hand, self-determination theory suggests that curricular and classroom activities should support the innate psychological needs for competence and autonomy to facilitate intrinsic motivation (Ryan & Deci, 2000a; Vansteenkiste et al., 2006). Various features of the instructional approach outlined above, such as the principle of voluntary participation, can be viewed as being in line with this guideline. On the other hand, however, many, if not most, educational activities are initially prescribed by the instructors or the curriculum and thus "a central question concerns how to motivate students to value and self-regulate such activities, and, without external pressure, to carry them out on their own"(Ryan & Deci, 2000a, p. 60)..

It is here where the *PROCON* curriculum management is especially vigilant, for example, in terms of cautiously and carefully terming and communicating to the students tasks, student obligations and course requirements, allowing for high degrees of freedom in self-initiation and choice. The use of pressure to coerce students is viewed as something to avoid or limit to a minimum; where needed, detailed and meaningful rationales and guidelines are issued to prepare students to help them make informed decisions and frame goals for learning in intrinsic ways (cf. Vansteenkiste et al., 2006).

In sum, the curriculum management actively reflects and makes an effort to design and influence processes of acculturation and the sociocultural (pre-)conditions of (individual) learning and the learning environment. By contrast, the attention paid to the interaction between students is more or less constrained to the in-class segments. In fact, attendance can be thought of as the main criterion or only 'mandatory' element for continued participation in the *PROCON* curriculum. (In the course of this study, the completion of questionnaires and tests had become the second necessary condition for certification.) This principle goes so far as to abstain from deciding who will or will not participate in the training course workshops for second-year student tutors (and, thus, organize and lead a first-year module as a student tutor).

Thus, in sum, the basic approach of *PROCON* can, in fact, be described as essentially committed to the "learning paradigm" (Barr & Tagg, 1995) of higher education. This is despite the managerial efforts to shape acculturation (and, thus, influence learning processes). In a similar way, the organizing control exerted by the programme management (e.g. when changing and adjusting modules included in the curriculum) and the mandatory elements described above are among the few aspects that can be viewed as linked to the more traditional "instruction paradigm" in the *PROCON* programme. Likewise, the *PROCON* management retains a more "instructionist" notion in its empirical focus on both process *and* outcome and its determination to shape an instructional system which aims at enhancing both the construction of knowledge and the progress skills in the learners. Nevertheless, as the methods as well as the procedures employed in the curriculum are generally and widely aligned to the core epistemological perspectives on learning and the learner outlined above, and one may justly speak of what Hannafin, Hannafin, Land, & Oliver (1997) called a “grounded approach” to the design of a learning system.

DIGEST 3.2

One key aim of the educational approach taken by the *Arbeitsgruppe "Beratung und Training"* (*BuT*) overseeing and coordinating the *PROCON* curriculum is to prepare students for professional practice in psychology through both the development of applied key skills and personal development. Two educational strategies promote the probabilities for achieving this and related objectives. On the one hand, *BuT* carefully intends to shape a supportive, resource-based, error-friendly, and motivating pedagogical context with high degrees of in self-initiation and choice to the participating learners, which both goal (e.g. Ames, 1990) and self-determination theorists (Ryan & Deci, 2000a; Vansteenkiste et al., 2006) advocate. On the other hand, the various methods and the overall instructional design of the core modules are based on Bandura's (1969, 1971, 1977, 1986, 1999, 2001) Social Learning Theory and its pedagogical principles (agency, enactive learning, modelling, observational learning).

3.3 Computer-Supported Learning in the *PROCON* curriculum

Demand and Initial Survey.

From the early beginnings of the *PROCON* curriculum in the mid-1990s, many participating students repeatedly voiced their desire to be able to observe models facilitating conflict resolution in groups and teams *before* they themselves were assigned mediation cases in the respective training module. This was also the result of an initial survey conducted after the *Conflict Resolution in Groups* module in cohort 2001/02 (see Figure 13 below). Here, 63 participants in cohort 2001/02 were asked to suggest in what ways the module or the curriculum could be improved. Nearly one quarter of all 120 individual statements received, and more than 40% of all responding students voiced their expectation that learning could be significantly improved if models or further preparatory support systems were offered (see Table11).

Table 11

Suggested Improvements of the 'Conflict Resolution in Groups' Module (based on an Initial Evaluation Survey conducted in 2001/02 prior to this study)

Cluster	No. of sugges-tions	Problem Descriptions/Solution Ideas (exemplary)	Percentage of total no. of sugges-tions	Percentage of total no. of students
Models/ Preparatory Learning Support Systems	26	• 'Conflict Resolution in Groups': Video or expert models needed; • Tutors' role model functions limited due to student status; • Materials needed to more fully understand strategy model described in the literature; • Module preparation remains too varied and uncontrolled; • Differences between two-party- mediation and multi-party group setting remain unclear	22%	41%
Case Analysis & In-Class Reflection Processes	26	• Role-based feed-back after simulation hard to formulate • Case analysis and reflection requires too much time • Methods and Tools needed to support case analysis and feed-back • Variety of methods for reflection too limited	22%	41%
Strategy Experiences	22	• First, 'initial agreement' stage cannot be authentically simulated in class despite its significance • Limiting simulation to one or two cases allows for experiencing the problems of transition between the stages of the strategy but obstructs variability of cases and case experiences	18%	35%
Class Times/ Time Allocation	22	• Constraints in total time; more time required for exercise • Time allocation: problematic especially on second day (concentration problems); shorter days possibly advantageous	18%	35%
Case Descriptions/ Case Simulation	16	• More diligent and careful preparation of role-players required by student tutors to allow for more authentic and free simulation • Case studies sometimes lack clarity and/or elaborateness • Some case simulations perceived as too complex and demanding for facilitating/mediating tandem	13%	25%
Other	8		7%	13%

Note. Based on 120 suggestions stated by 63 participants. Percentages rounded.

Various reasons may account for the fact that this idea was adopted only recently.

To begin with, the Conflict Moderation stage model as it was introduced into higher education in the mid-1990s (Redlich, 1995) allowed for significant degrees of inventiveness in the facilitative methods applied at the various stages. Moreover, the newly invented *act-storming* technique proved a productive resource for the inventive generation of alternative third-party behaviours. Unsurprisingly, concerns surfaced in regard to these creative processes possibly being affected by a premature provision of models – even intensified by the prospect of the models reaching an exemplary status which could later prove hard to correct.

Furthermore, the traditional and predominant method of delivery (in both higher education and the Curriculum) had been was face-to-face instruction. At the time, the *BuT* curriculum designers had had limited experience with other methods of delivery such as video-based lecturing. Then, producing and presenting models of conflict resolution facilitation in line with both the pedagogical and theoretical approach implied a recourse to additional live instruction, preferably in form of a supplementary module to be implemented prior to the *Facilitating Conflict Resolution in Groups* training course. However, this implication seemed unfeasible in view of both financial constraints and the probable need to replicate such an additional module due to innovations in technique and coherence.

Development and Implementation Process.

Meanwhile, experiences were made with the production and use of computer-based software for facilitation (Nitor GmbH, 2004) and communication skills in counselling (Jacobs, 1997; Jacobs & Redlich, 1998) as well as web-based training programmes for soft skills (Nickels, Redlich, & Tendler, 2003; 2002; Bildat, 2003).

The resulting hyperware products were essentially based on observational learning and situational judgement and thus suggested compatibility to the *PROCON* instructional approach. Finally, in 2002, additional grants for multimedia production and development in higher education were made available through the municipal *Behörde für Wissenschaft und Gesundheit* (Department of Science and Health) of the City of Hamburg, Germany. A successful application by the *BuT* Arbeitsgruppe supplied funding needed for the development of multimedia software intended to be used in the *PROCON* curriculum to improve student preparation for the *Facilitating Conflict Resolution in Groups* training course (see below). Moreover, it was hoped that, if the software was tailored carefully, it would also lead to some improvements in related clusters of Table 11. For example, it was hoped that software preparation could also promote the development of a common language by providing a

taxonomy and models for strategy and feed-back. Thus, it could possibly ease feed-back processes, common case analysis, and diminish the need for experiencing all stages and stage transitions during simulation.

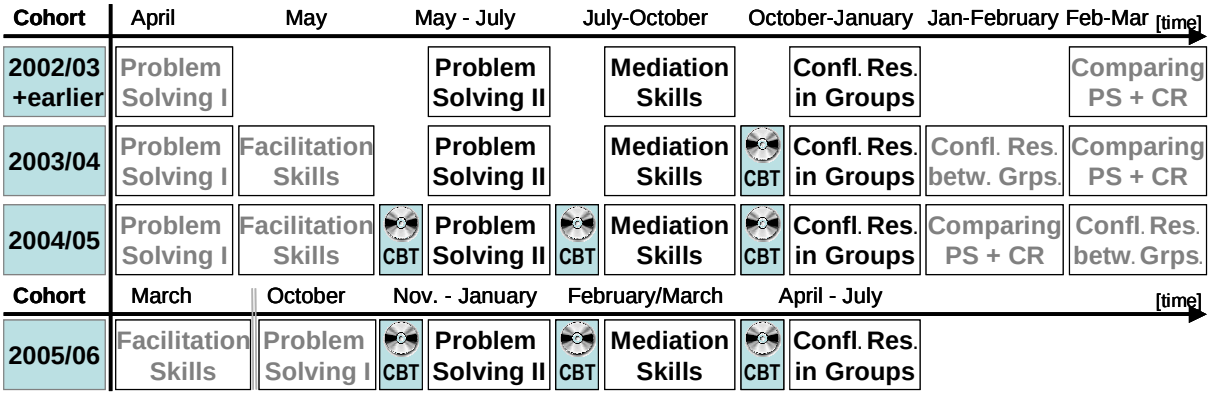


Figure 13. Implementation of technology-based preparation modules into the *PROCON* Curriculum.

Starting in 2003, computer-based software (Nitor GmbH, 2004) was offered to students to accompany their preparation for the *Facilitating Conflict Resolution in Groups* training course (which, however, is not part of the core curriculum depicted in Figure 12; see section 3.1). In terms of both spin-off and original research products, two additional computer-based software programmes (Kilburg, 2005a; ProKonflikt Team, 2004b, 2006b) were developed subsequently to intensify preparation for the two other training modules (see Figure 13 for an overview of the implementation process). By 2004, all training courses were delivered in a "blended learning" format.

Approach to Blended Delivery.

All computer-based software modules used in the core curriculum (which will be referred to as 'courseware' in the following) were intended to enrich and deepen preparation previously aided by studying extant written materials, mainly books (Klebert, Schrader, & Straub, 1997; Redlich, 2004a; Spitzer & Evans, 1997; Thomann, 1998). To this objective, (conceptual) knowledge about the basic behavioural strategies was not extensively imparted, neither in the *Problem Solving* nor the *Conflict Resolution* coursewares; rather, students were advised to revert to the written material in order to create a knowledge base and prepare themselves for or intensify learning with the respective courseware.

This approach to implementing courseware had little impact on the design of the subsequent training courses as they continued to focus on what is probably best achieved in purely synchronous, face-to-face, resident classroom situations, namely, social learning through

enactive and reflective problem-based tasks by means of role-play, subsequent feed-back and interactive collaborative group discussions. The resulting blend of traditional and computer-supported delivery modes can also be conceptualized in terms of the expanded 3C model discussed above (see section 2.5):

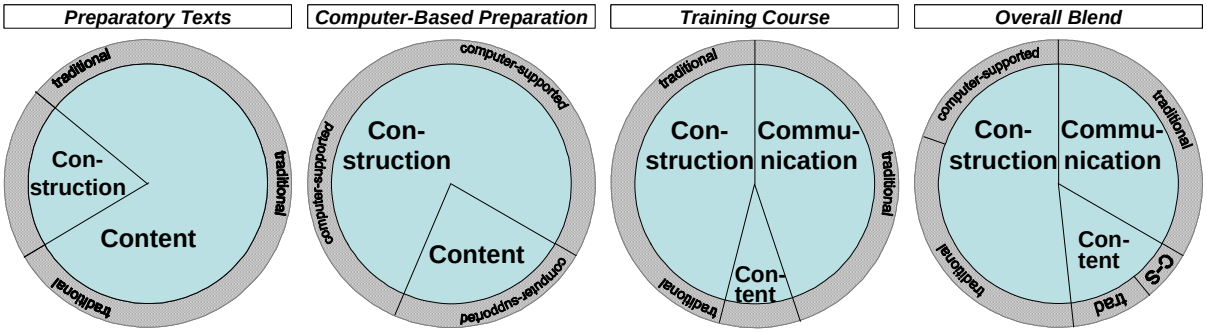


Figure 14. Extended *3C Model* conceptualisation of the typical delivery mix in PROCON core modules.

As preparatory or accompanying written materials are seen as focusing on presenting factual knowledge, the coursewares are viewed as aiding solitary processes of knowledge construction of social situations, and classroom interaction mainly focused on fostering communication and both individual and collaborative knowledge construction. This conceptualisation is summarized in Figure 14.

Courseware Design and Pedagogy.

Matchware® Mediator Expert (versions 7 and 8) were used to create the coursewares. These were exclusively distributed on single CD-ROM disks 4 (or more) weeks prior to the respective training courses. Students were given accompanying written instructions on how to install and use the coursewares on either their own personal computer or the Department of Psychology's computer facilities.

In terms of design and layout, the coursewares can be characterized as guided, sequential tutorials (Schulmeister, 2002a), as the design of all coursewares was largely based on a linear story which represents the chronological or stage order in which problem solving and conflict resolution is thought to take place. Thus, as in respective real situations, knowledge and a conceptualization of previous events is often necessary to understand present situations and tasks and to generate appropriate mediation behaviours. In turn, users were invited to follow a defined learning path, albeit cross-chapter and cross-task user navigation was made possible (see Figure 15), supported by process indicators which signal visited pages and completed tasks. Therefore, users were largely free to use only parts of the courseware or

to select alternative learning paths. Kilburg (2005b), drawing on Schulmeister's (2002b) suggested classification of interactivity, rates these types of software as moderately to highly interactive (grade II to III).

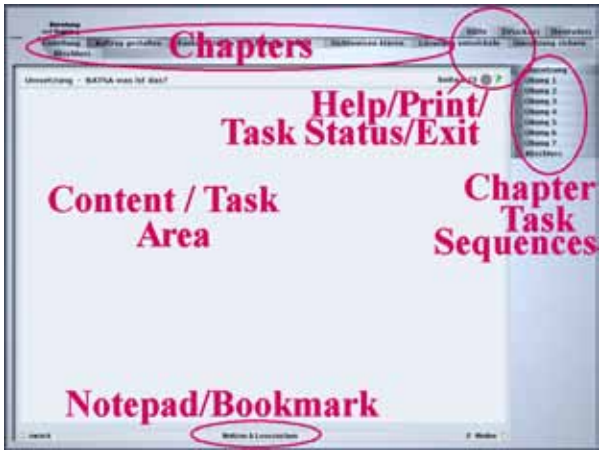


Figure 15. **Typical navigation and toolbars in *BuT* coursewares (based on screenshot taken from *Conflict Resolution in Groups* courseware, ProKonflikt Team. 2006).**

The interactive coursewares contained a series of video scenes intended to recount a whole story narrative to demonstrate how a theoretical approach or strategy is put into practice. For educational purposes, the storyline was broken up into chains of events in line with the underlying strategy. Subsequently, and educational video scenes were produced in order to capture major events or to demonstrate major difficulties and demands in the inter-personal interaction. One sequence template frequently drawn on in the various coursewares (and most often used especially in the computer-based preparation softwares in both the *Mediation Skills* and *Conflict Resolution in Groups* coursewares) consisted of the following series of tasks (see Table 12 below):

(1) Instructions and Questions

The learners were asked first to carefully read the written instructions presented by the software. The instructions usually detail the setting, history and incidents of a scene in which a mediator or facilitator is confronted with a specific demand characteristic in the on-going inter-personal interaction. Students were also instructed to identify with the mediator or facilitator and to prepare themselves to respond to either the situation presented (e.g. "What would you say or do?") or to a specific question related to the underlying strategy (e.g. "How would you word your suggestions to proceed?").

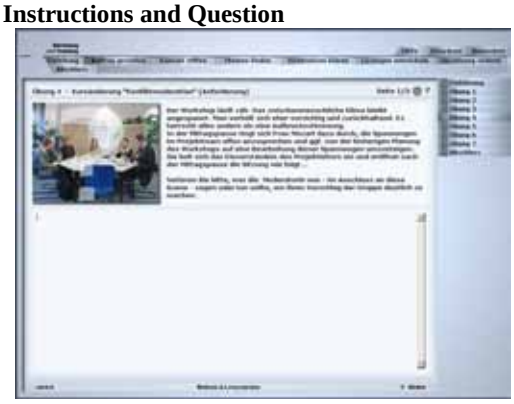
(2) Video Sequence (demand situation)

Students were then asked to watch the high-fidelity video sequence and to detect the various demand characteristics embedded in the situation. The duration of the various sequences vary between 15 and 150 seconds; however, all sequences were carefully checked and cut to a maximum length in order to avoid over-burdening of the learners due to an overwhelming situational complexity. Learners were given the possibility to re-play and watch the individual scenes repeatedly. In the various revisions of the software, the learner's control of the video player (pause, rewind, forward, etc.) was gradually enhanced.

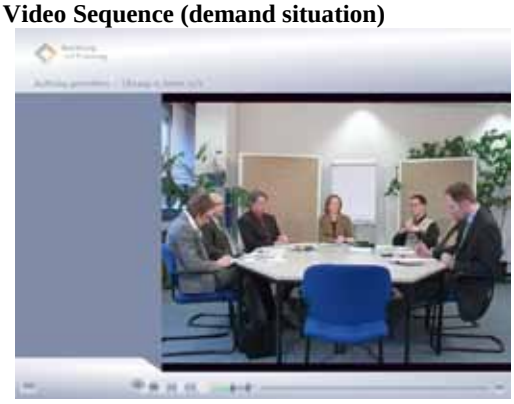
Table 12

Paradigmatic Learning Sequence (Template used in BuT *Computer-Based Preparation Modules)*

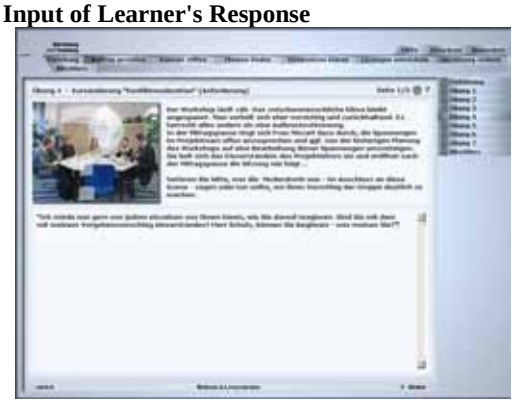
1) Instructions and Question



2) Video Sequence (demand situation)

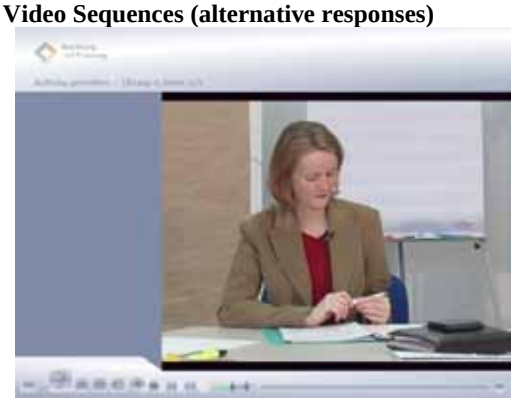


3) Input of Learner's Response



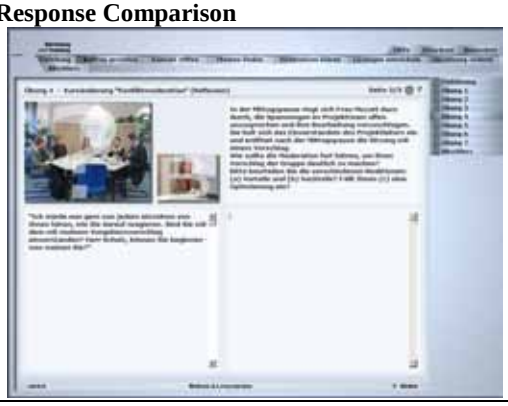
4) Presentation of Alternative Responses

5) Video Sequences (alternative responses)



6) Response Selection & Feedback Provision

7) Response Comparison



8) Reflection Task

Note. Screenshots taken from *Conflict Resolution in Groups* courseware (ProKonflikt Team. 2006).

(3) Input of Learner's Response

By use of the keyboard, students were then prompted to input their responses to the question posed in step (1). In most cases, students were asked to type their response in *oratio directa* to promote identification of the learner with the video character and to avoid and indirect discourse, i.e. having students *describe* what they would do or say.

(4) Presentation of Alternative Responses

Subsequent to the learner's input, students were invited to watch three video sequences showing the protagonist (i.e. the mediator or facilitator) perform alternative responses to the demand situation. This elaborates on Gentry's original approach.

Learners were given the task to rate the alternative responses in terms of their "appropriateness", defined as

- adherence to the underlying theoretical model or strategy (as described by the authors of the base literature and other written materials) and
- the respective fit to and inclusion of situational demand characteristics.

(Although this may be a viable alternative use of this learning sequence, students were not asked to rate the responses in terms of their proximity to their own response or as to the degree of congruence with their internal contingencies, i.e. their cognitive-affective situation or their 'inner team' as described in section 2.2 above.)

(5) Video Sequences (alternative responses)

Then, learners were asked to watch the video sequences randomly marked as A, B, and C. In order to arrive at a rating of the video sequences, students could also re-play the demand situation video sequence. The three alternatives had been designed in such a way that, at first glance, they often seemed similar - in form, duration and, at the outset, even in content. Often, two of the alternatives showed responses that could be viewed as 'typical' or 'common' action impulses; a third, more differentiated response provided what can be thought of (in Piagetian terms) as a "dosed discrepancy" to affect the learner's equilibrium and initiate changes in cognitive-behavioural schemas.

However, it was verified that experts assisting in the design of the courseware (i.e. authors, former students, colleagues, and reviewers of the accompanying written materials on which the courseware was based) found a relatively dependable answer as to which alternative was most appropriate (in terms of the definition outlined above).

(6) Response Selection & Feedback Provision

Learners were subsequently expected to mark the most appropriate alternative and were then provided with feedback to each alternative. The feedback regularly consisted of expert comments to features of the underlying theoretical model or strategy, to the situational demand characteristics (as conceptualised by the experts) and to possible effects of the mediator's or facilitator's actions. Possible shortcomings of an alternative were also frequently discussed, even when the alternative was rated most appropriate.

(7) Response Comparison

The learners were then given access to the video sequences of the original demand situation and the most appropriate alternative, and their initial formulation of a possible response. Bearing in mind the feed-back received previously, they were encouraged to review their own initial response and, finally, compare it to the most appropriate alternative offered.

(8) Reflection Task

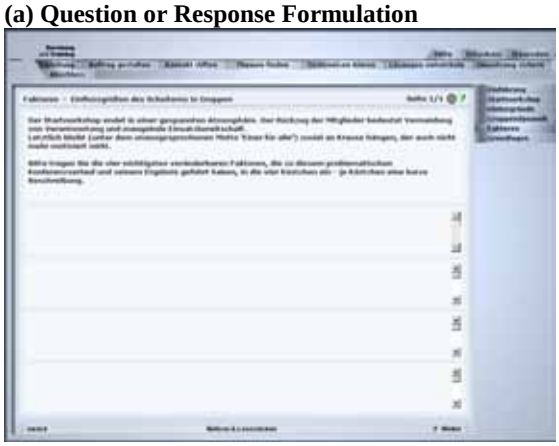
The final task of the template often consisted of a question intended to foster reflection on the experiences made, i.e. taking into consideration what they have newly discovered as to features of the underlying objectives or the situational demands, or as to its possible effects. To achieve this objective, four main variants of the reflection task were used most often. Either, learners were asked to improve the wording of their own suggested response. Alternatively, students were encouraged to re-write the presented, most appropriate response to arrive at an even improved fit. A third variant of the task targeted at having students reflect on the degree as to which the most appropriate response would differ from a response that is in congruence with their personality and internal cognitive-emotional situation. Or, finally, students were asked simply to list advantages and disadvantages of their own (or the most appropriate) response.

This sequence has been described as 'paradigmatic' for the individually focused computer-supported cevelopment of social skills by Rogmann and Redlich (2007, see section 2.5). Variations of this sequence were also used, for example, with an exclusion of the learner's initial response (step 3) or the comparison and reflection tasks (steps 7 and 8).

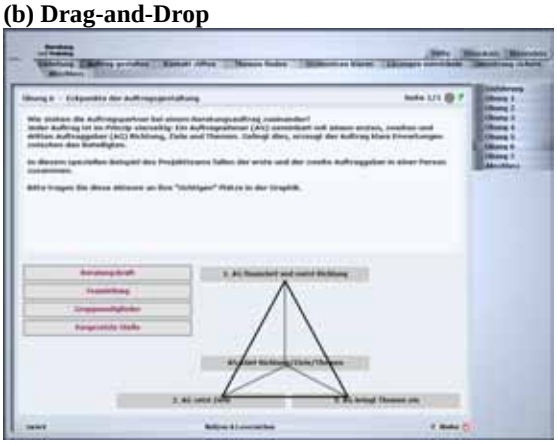
Table 13

Other Task Templates Used in BuT *Coursewares*

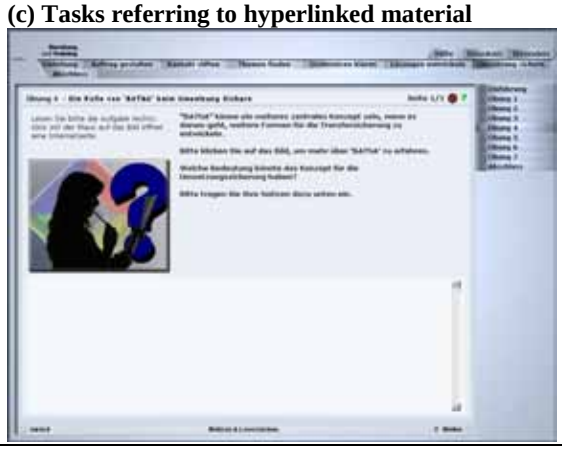
(a) Question or Response Formulation



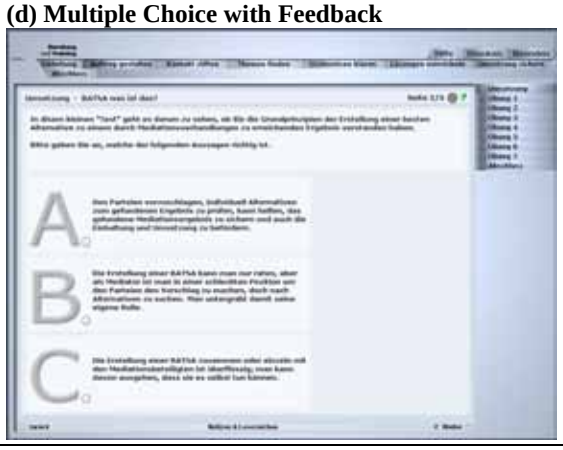
(b) Drag-and-Drop



(c) Tasks referring to hyperlinked material



(d) Multiple Choice with Feedback



Note. Screenshots from *Conflict Resolution in Groups* courseware (ProKonflikt Team. 2004; in 2006 layout)

Much less frequently, other task templates, such as drag-and-drop exercises, written tasks or feedback based on multiple choice questions were employed (see Table 13). Most of these tasks and feedbacks were designed to support learners in their own individual knowledge construction. For example, much in line with the SQ4R study method (Robinson 1970), learners were prompted at the introduction of chapters to formulate their own questions in respect to the content to be presented. These were re-displayed at the end of a chapter for response and refinement.All inputs of the learners were recorded and stored. In later versions of the coursewares, learners could print out all tasks, inputs, and feed-backs. Unfortunately, however, in the first editions of the coursewares, the way the information was stored, the accessibility for users, the options to transfer the information from the leaner's computer to tutors, other learners, programme management, or evaluation systems as well as the analysis of the information received all proved extremely complex, limiting an effective and valid analysis; it was therefore decided not the include these data in this study.

Content.

As indicated above, the *BuT* coursewares are typically based on linear storylines representing the chronological stages of the underlying problem solving and conflict resolution models. Each courseware takes approximately 6-10 hours to complete.

The *Problem Solving II* courseware developed by Kilburg (2005b, please see here for further details) consists of series of videos in which a facilitator assists a task force group in a medium-sized grain mill corporation appointed to tackle the problem of contaminated flour products. Based on this case, the courseware videos show the facilitator carrying out problem analyses, a root cause analysis, the generation of feasible solution scenarios, a decision-making procedure, and an analysis of potential problems and opportunities.

Within a research and development programme funded by the Hamburg Ministry of Science and the state government, the six-university E-learning Consortium Hamburg (ELCH) decided to grant financial support needed for the development of the *Conflict Resolution in Groups* courseware (ProKonflikt Team, 2003, 2004a, 2006a) between 2002 and 2004. Here, a scenario is staged in a medium-sized corporate supplier for hospital goods; the corporation provides appliances and articles for inpatient care (e.g. injection and infusion devices, bandages, linen, etc.). A 7-member strong project group is given the task to suggest a combined marketing-, supply- and sales strategy in order to broaden the client base (and possibly include smaller enterprises, such as home care providers). The learners are first

introduced to the various parties involved (and their respective views) before they see the project team in action and, indeed, in conflict and are asked to analyse problematic psychosocial aspects. Then, they are asked to identify with an external mediator invited by the team leader to facilitate group work. The video scenes embedded in this courseware cover how the mediator guides the team though the prototypical stages of team mediation as described by Redlich (2004a, 2006; Rogmann & Redlich, submitted for publication): contracting, identifying conflicting issues and reaching an agreement on conflict resolution procedure, clarifying perspectives and underlying interests, cultivating positional shift and finding agreements, safeguarding transfer and implementation.

Despite being used earlier in the curriculum, the *Mediation* courseware has been developed chronologically after the *Conflict Resolution in Groups* courseware and, in fact, further exploits a one-two-one conflict situation that surfaced in the project team scenario. The two respective parties agreed not to undergo meditation in a group setting, but rather conduct a confidential series of sessions with the same mediator. In this courseware, the educational focus is placed not so much on the stages of mediation; the intention rather was to focus on basic competencies required for mediation, such as structuring the mediation process, active listening to two sides to explore underlying goals, needs, and emotions, creating and maintaining a fair atmosphere, and generating options. Moreover, accompanying mediation techniques are imparted, such as psychodramatic doubling for mediation (cf. Thomann & Schulz von Thun, 1988).

DIGEST 3.3

As a majority of participating students repeatedly voiced their desire to be able to observe expert models while facilitating conflict resolution in groups and teams, coursewares for the core modules of the *PROCON* curriculum were gradually developed and introduced between 2002 and 2006. The interactive coursewares, designed to enrich and deepen module preparation previously aided by studying existing written materials, contained a series of video scenes intended to recount a whole story narrative to demonstrate how a theoretical approach or strategy is put into practice. For educational purposes, video scenes were produced end embedded in software to achieve a higher degree of interactivity. One learning sequence template frequently drawn on in the various coursewares consisted of the presentation and development of a response to a given demand situation, the observation and rating of alternative responses, the analysis of feedback and an extended reflective task.

3.4 Evaluation Research Design

All evaluative approaches taking account of outcome or impact parameters have one common element. If the merit or the value of a programme is to be determined, a standard is needed against which the found characteristics or outcomes of the programme can be compared. As Mohr (1995) put it:

> "The Crux of the analysis of the efficacy of a treatment or program with respect to a particular outcome dimension Y is *a comparison of what did appear after implementing the program with what would have appeared had the program not been implemented.* Events in the what-would-have-happened category must obviously be troublesome. … This element can never be observed and can never be known for certain. Its paramount importance in the assessment of the efficacy of a program and, at the same time, its fundamental inaccessibility make this the pivotal point of all impact analysis designs and a major source of reservations about the validity of evaluative conclusions. We will refer to this element as *the counterfactual* … .' (p. 4)

Estimating this "counterfactual" is at the heart of evaluation designs; be it the employment of control groups in a controlled experiment design, the use of comparison groups in quasi-experimental settings, the utilization of pre-tests in one-group designs or repeated testing strategies in within-subject or single-n designs. Statistical procedures are used to quantify the comparison of the programme outcome to the estimated counterfactual; be it the analysis of the difference of group means or proportions or the analysis of regression coefficients. In the following, the evaluation design tailored to the specific setting and pedagogies, the issues and assumptions under review and the very conditions of evaluation will be detailed.

Integrative Evaluation Strategy.

In a series of interviews conducted with programme management during the planning stage, both major conditions for evaluation and outcomes of interest were identified. In line with the experiences reported by Draper et al. (1996), it became clear that programme management was interested in determining a summative impact of both the various modules and the introduction of computer-supported elements as well as obtaining ideas for changes through formative analyses. With formative feed-back regularly provided during the course of the evaluation analyses and the nature of this report being a doctoral dissertation (rather than a final evaluation report), the summative aspects are focused on in the following. (The evaluation guidelines detailed in Chapter 2.8 also chiefly address impact-related questions.)

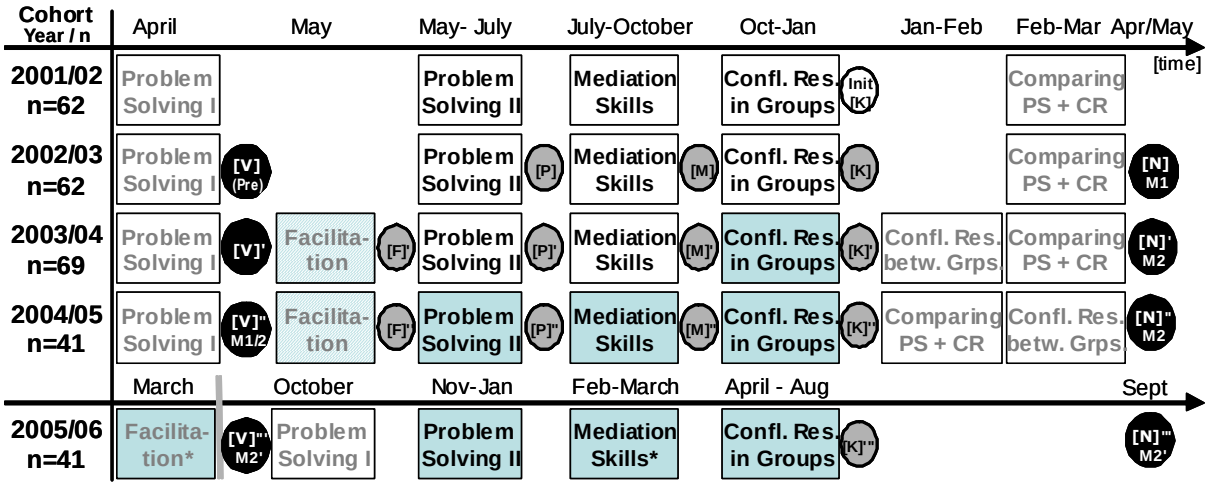
In order to identify outcome and impact parameters for course packages and the overall programme, it was decided to implement a two-fold evaluation design within the existing programme. Despite the widespread and critique, the "gold standard" for both traditional social research and impact evaluation still is the true experiment entailing a near-perfect control group and randomized errors. It seemed reasonable, therefore, to implement an evaluation design that is oriented towards this ideal goal. At the same time, however, it was necessary to follow and observe a number of practice restrictions. For example, programme management made it a condition for evaluative research that the students' traditional high degrees of freedom in self-initiation and choice thought to contribute to motivation were not undermined by research efforts. Thus, "external pressure" and the notion of control always inherent in evaluation instruments had to be reduced to an utmost minimum, and randomization was not an option. Moreover, as all interested students of one student cohort were given permission by programme management to participate, it seemed both uneconomical and futile to try to implement control groups by using students other than those participating.

Two additional conditions established initially by programme management can be considered typical for higher education practice settings. However, they also further complicated the implementation of a quasi- experimental and/or reflexive control design strategy. First, due to both legal and cultural concerns, the content of pre-tests before 2004 had to be restricted to biographical and demographical variables and such student characteristics either considered as required or necessary at programme entry. Thus, straightforward reflexive pre-post-comparisons were rendered infeasible. And, second, programme management also cautioned evaluators that, due to financial restrictions, the unpredictability of peer-tutor availability/motivation and the various exchanges of experiences between programme management, teaching staff and peer tutors, the content and the module system was subject to repeated change. Thus, a flexible evaluation system was required with likely amendments of the instruments used.

In sum, measures intended to improve equivalence of research groups and/or to reduce selection bias was therefore beyond the control of evaluation. As a result, (a) the multiplicity of module groups within a cohort and (b) entire consecutive student cohorts with differing degrees of computer support were identified as groups most proximate to quasi controls.

Design.

Starting with the 2002/03 cohort, it was communicated to all students voicing interest in participating in the *PROCON* curriculum that the project was not simply intended to offer additional opportunities for meaningful learning but also, in exchange, to generate opportunities for educational research. Consequently, the programme was re-labelled, adding "Lehr , Lern- und Forschungsprojekt" ('teaching, learning, and research project') to its original title.



Cohorts/Modules:
□ – Training module (classroom/peer-tutored), ▭ – Blended learning module (custom-made courseware, obligatory),
▭ – Blended learning module (off-the-shelf commerical courseware, voluntary), * – Module change (see text),
n – Number of participants who completed Pretest (2001/02: Init CR) in cohort year and Posttest within study duration

Instruments:
● – Written examination, ◎ – Post-module online survey/questionnaire, ○ – Post-module written survey/questionnaire,
[V]– Pretest, [N]– Posttest, M1– includes multimedia-test (Form 1), M2– incl. multimedia-test (Form 2)
Init– Initial Evaluation Survey, ' / " / '" - instrument alterations/design number
Post-Module Surveys/Questionnaires: [F] – *Facilitation Skills* module, [P] – *Problem Solving* module(s), [M] – *Mediation Skills* module, [K] – *Conflict Resolution in Groups* Module

Figure 16. Temporary module and evaluation design pattern.

Figure 16 above depicts the temporary module and evaluation design pattern from a retrospective point of view. After each of the core modules, a tailored online survey (◎) was conducted across the various tutored module groups, containing questions targeted at the use of computer-supported elements, the self-reported increase of knowledge and skills, and the participant's confidence in both theoretical understanding and conflict resolution/mediation skills.

Under the conditions set out above, written pre- and post-tests (●) were developed and re-adapted annually. Pre-tests were to determine pre-requisite knowledge and skills considered necessary at programme entry (such as knowledge about 1-to-1-counselling and group facilitation techniques) whereas post-tests contained both knowledge tests as well as video-based situational judgement tests to determine knowledge and skill level at the end of

programme participation. In 2004, programme management allowed for the introduction of post-test elements in pre-test which made the implementation of a reflexive control design essentially possible.

Moreover, to determine the impact of the introduction and learner's use, it was also decided to run ex-post facto analyses by dividing the entire sample according to the actual utilization of computer-supported learning elements by students. This evaluation strategy allows for ex-post comparisons between post-module surveys and post-tests. Here, however, changes in instrumentation will have to be accounted for.

Online surveys were regularly conducted between modules. Much in line with the action research paradigm, summaries of online survey results were feed-backed to both programme management and student tutors to both probe for validities and to inform these important stakeholders about areas possibly in need for improvements. Subsequently, the results of the online surveys will be included as they may provide essential information about variables possibly mediating the impact of courseware introduction and use. Moreover, it was hoped that repeated measures could give indications about customary learning processes or trends in terms of chronological learning gains and changes in self-reported self-efficacy.

DIGEST 3.4

The overall evaluation strategy followed in this study had to take into consideration various practice restrictions often found in tertiary education settings which essentially rendered randomized or quasi-experimental designs both impossible and futile. In turn, a flexible evaluation system within the existing programme was developed which identified both the multiplicity of module groups within a cohort and entire consecutive student cohorts with differing degrees of computer support as groups most proximate to quasi controls. Under the given conditions, written pre- and post-tests as well as module online surveys were developed and re-adapted annually.

3.5 Participants

In total, 272 different students took part in one or more modules offered by the *PROCON* programme between 2002 and 2006 and completed one or more of the instruments. Due to the nature of multimedia-supported written tests and online-surveys, the data of three students with disabilities (hearing deficiency / visual impairment) were excluded from further analysis, resulting in 268 students completing one or more instruments during the duration of this study.

As mentioned above, students were allowed to suspend their participation in the programme after completion of a module (and the respective module instrument) and re-enter the programme at the same stage of a subsequent cohort. Thus, the sample size for each evaluation instrument administered in each cohort was subject to continuous change. The following Figure 17 shows a summary of the resulting sample sizes, attrition and migration rates for all cohorts and instruments included in this study.

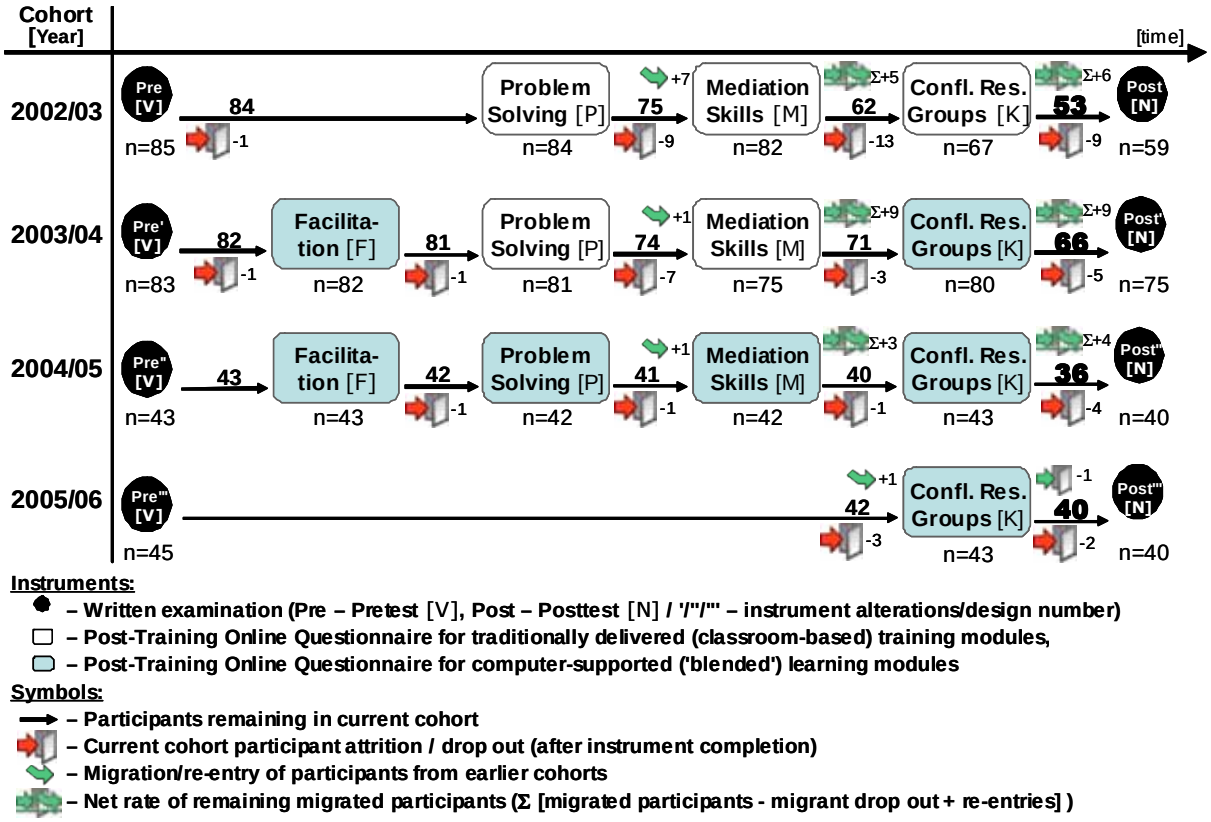


Figure 17. Cohort sample sizes, attrition and migration rates.

For example, 85 students were pre-tested in 2002; of these, 84 completed the *Problem Solving* modules, and 75 of these completed the *Mediation Skills* module. However, as 7 additional students were allowed to re-enter the curriculum at this point, n=(75+7=)82 completed the respective module online questionnaire. Finally, keeping migrant and original students separate, 59 students completed the post-test in 2002/03 of which 53 were pre-tested in the same cohort and the remaining 6 were migrated from earlier cohorts/modules.

For various reasons, the evaluation analyses reported below focus on full cohorts (i.e. without migrating students). To begin with, the annual changes in technology-supported elements are best evaluated when comparing groups exposed to these changes. Students re-entering the curriculum from earlier cohorts likely have a different learning background. Furthermore, both the knowledge and performance status of re-joining students as well as their self-rated confidence in the application of necessary skills are likely to differ from regular partakers due to the long recess. The same holds true for students had suspended their participation before 2002 and re-entered at some point during study duration. For these 12 participants, it seemed both impractical and uneconomical to have a pre-test (in which information on various background variables was collected).

The following Table 14 gives socio-demographic details on the remaining (268-12=) 256 participants, and the differences between the sample base and those excluded from most of the analyses reported below. Examining the equivalence principle, differences were tested for statistical significance; the results are also reported in Table 14. Unless notes otherwise, all calculations reported below were performed using the Windows 11.5.1 version of the SPSS® software. No significant differences were found between the cohorts sample bases and those not included in most of the subsequent analyses. As shown in Table 14, the percentages and means of most socio-demographic variables seem similar.

With minor exceptions, the analyses performed for the various individual cohorts yields the same results; it and can be found in Appendix D.1. In sum, at this stage of the analysis, the data give no reason to suspect that the individuals excluded from the subsequent analyses may not have been equivalent to those included in the further analyses.

Table 14

Pre-test Equivalence of Sample Base and Excluded Participants: An Ex-Post Facto Analysis

Variable	Sample Base	Excluded	Equiv. Tests
Participants (n)	**195**	**61**	
Age (yrs)	29.8 (6.4)	29.3 (6.1)	U=5627
Sex (% female)	80.5%	83.6%)	χ^2=.29
Duration of study (semesters)	7.8 (2.6)	8.5 (3.3)	U=4855.5
Level of field experience:			
(% with exp. in 1-to-1-counselling)	58.3%	56.7%	χ^2=.05
(% with exp. in facilitating)	57.2%	62.3%	χ^2=.49
Occupational experience:			
(% with vocational education)	50.5%	46.7%	χ^2=.27
(total number of years)	6.5 (5.5)	7.5 (5.2)	U=3971.5
(% w concurrent part-time-job)	74.4%	78.7%	χ^2=.47
(h/week)	10.9 (8.9)	11.9 (7.5)	U=5408.5
Personality variables (BIP, 2002-04): n	155	56	
Kontaktfähigkeit (Contact skills)	66.6 (12.0)	69.8 (8.9)	U=3779
Führungsmotiv. (Motivation to Lead)	53.8 (10.9)	54.0 (10.4)	U=4336
Sensitivität (Sensitivity)	52.9 (6.1)	52.4 (5.6)	U=4042.5
Selbstbewußtsein (Self-Confidence)	54.4 (10.0)	55.0 (9.9)	U=4157.5
Conflict Styles (DUTCH-D):			
Yielding (Nachgeben/Einlenken)	12.0 (1.5)	12.0 (1.6)	U=5879
Problem Solving (Problemlösen)	15.9 (2.2)	15.9 (2.4)	U=5833
Compromising (Kompromisssuche)	15.2 (2.3)	15.1 (2.7)	U=5668.5
Avoiding (Vermeiden)	11.1 (2.6)	10.6 (2.2)	U=5501.5
Forcing (Durchsetzen)	10.9 (2.3)	10.9 (2.4)	U=5828.5

Note. Data are arithmetic mean (standard deviation) unless stated otherwise. Mann-Whitney-U tests were calculated as application prerequisites for Student t-tests (df=N_{ex1}+N_{ex2}-2) were not met (insignificant results of both the Kolmogorov-Smirnov-Test with Lilliefors-correction and the Shapiro-Wilk test, insignificant Levene test). χ^2 values (df=1) are reported in case all cell frequencies > 5. Two-tailed equivalence was assumed for cohorts with *p<5% and **p<1% probability levels.

For purposes of later inter-cohort-comparisons, equivalence of cohorts comprised of the sample base is an important prerequisite. Therefore, the same background variables required were also tested for differences across cohorts. The results reported in Table 15 below indicate that the cohorts may differ in at least three respects.

First, the portion of participants with previous field experience in group-based facilitating differs significantly across between cohorts. In particular, it is notably smaller in the 2005 cohort. Second, the participants' self-reports in terms of their temporal occupational experience was found to differ markedly between cohorts. While, for example, the mean occupational experience (in number of years in vocational jobs) is highest in the 2005/06 cohort (M=7.7), it approximates only half that amount (M=3.9) in the 2004/05 cohort. And third, one personality variable has also been found to yield a significant difference between the three cohorts 2002 to 2004. This is probably due to indicated differences between the 2002 and 2003 cohort (Bonferroni-corrected t=2.97, p<5%, two-tailed).

Table 15

Equivalence of Participant Cohorts (Sample Base Comparison)

Variable Cohort	2002/03	2003/04	2004/05	2005/06	Equiv. Tests	p
Participants (n)	**53**	**66**	**36**	**40**		
Age (years)	31.4 (7.3)	29.1 (5.1)	29.2 (6.3)	29.5 (6.8)	χ^2=3.4	
Sex (% female)	77.4%	83.3%	77.8%	82.5%	χ^2=0.9	
Duration of study (semesters)	8.7 (3.8)	7.6 (2.0)	7.2 (2.1)	7.5 (1.7)	χ^2=3.9	
Level of field experience:						
(% with exp. in 1-to-1-counselling)	60.8%	60.0%	58.3%	52.5%	χ^2=0.8	
(% with exp. in facilitating)	66.0%	63.6%	68.6%	22.5%	χ^2=27.5	***
Occupational experience:						
(% with vocational education)	58.5%	45.5%	42.9%	55.0%	χ^2=3.2	
(total number of years)	6.5 (6.2)	5.3 (4.9)	3.9 (4.4)	7.2 (6.0)	F(3,191)=2.7	*
(% w concurrent p/t-job)	81.1%	72.7%	61.1%	80.0%	χ^2=5.3	
(h/week)	14.6 (9.2)	13.9 (6.7)	11.2 (5.5)	13.2 (7.6)	χ^2=1.8	
Personality variables (BIP, 2002-04):						
Kontaktfähigkeit (Contact skills)	64.6 (10.6)	67.4 (12.4)	68.1 (10.6)		F(2,152)=1.2	
Führungsmotiv. (Motivation to Lead)	53.2 (10.4)	54.4 (11.4)	53.6 (10.4)		F(2,152)=0.2	
Sensitivität (Sensitivity)	51.3 (5.8)	54.3 (6.2)	52.8 (5.9)		F(2,152)=3.6	*
Selbstbewußtsein (Self-Confidence)	54.2 (9.2)	55.1 (10.1)	53.4 (11.2)		F(2,152)=0.3	
Conflict Styles (DUTCH-D):						
Yielding (Nachgeben/Einlenken)	11.8 (1.6)	12.0 (1.4)	12.4 (1.5)	12.0 (1.3)	F(3,191)=1.4	
Problem Solving (Problemlösen)	16.1 (2.3)	15.9 (2.1)	16.6 (2.0)	15.2 (2.4)	F(3,191)=1.3	
Compromising (Kompromisssuche)	15.0 (2.5)	15.6 (2.3)	15.4 (2.4)	14.8 (1.9)	F(3,191)=1.3	
Avoiding (Vermeiden)	10.8 (2.6)	11.0 (2.5)	11.2 (2.9)	11.7 (2.6)	F(3,191)=1.1	
Forcing (Durchsetzen)	10.5 (1.9)	10.9 (2.3)	11.1 (2.8)	11.1 (2.5)	F(3,191)=0.7	

Note. F($df_{between}$,df_{within}) values of one-way ANOVA were calculated for variables with interval level of measurement. Where the Levene-Test yielded significant results, χ^2 values were calculated using the Kruskal-Wallis H rank test (df=3). For nominal data, the results of χ^2 tests for the resulting contingency tables (df=3) are reported. Equivalence was assumed with *p<.05, **p<.01, and ***p<.001 probability level (two-tailed).

Whether these significant statistical differences can be taken as indicators of real-world phenomena, and whether they can be considered to influence or interact with the variables of interest in this study will be essentially a matter of further empirical investigation, and, accordingly, the subsequent analyses reported below will include examinations relevant to this issue. For the time being, however, the assumption that the various cohorts of the sample base are generally equivalent is upheld.

DIGEST 3.5

In total, 272 different students took part in one or more modules offered by the *PROCON* programme between 2002 and 2006 and completed one or more of the instruments. For various reasons detailed in this section, the evaluation analyses reported below will focus on those 195 participants completing the curriculum within one year and cohort. The results of statistical testing intended to detect differences between those included and those excluded from the sample base did not generally contradict the assumption of their pre-test equivalence.

3.6 Principles of Instrument Design and Implementation

Anonymity and Efficiency.

To coordinate registrations and to perform the necessary checks needed for a thorough organisation and administration of the curriculum, the personal details of participating students were required. On the other hand, however, students were encouraged to articulate their impressions and concerns in order to obtain realistic evaluations. To promote self-disclosure and minimize both possible reservations by students in regard to sensitive information and social desirability bias, a two-fold strategy was used in instrument implementation.

First, where possible, computer-administered forms were used for evaluations, ratings, surveys, and necessary assessments as computer-supported administration is generally thought to elicit more self-disclosure than traditional paper-pencil forms of administration or standardized interviewing (Weisband & Kiesler, 1996). (Computer-forms also had an efficiency advantage in that information was directly stored and transmitted in digital form.)

And, second, all information obtained from students was by use of an anonymizing, albeit individually reproducible code. This seven-character-long code consisted of a series of letters and figures, based on information only known to the individual participant and not to curriculum management. To produce and reproduce the code over time, a student had to go through a series of questions pertaining to relatively stable information, i.e. such that could possibly be taken from official documents. For example, the code included the first letter of the first name of one's father, the number of biological siblings one has been raised with until the age of 16, the last figure of the university registration (student ID) number, or the first letter of the place of birth (in German language). As the only link between the student's real name and his or her code (and, thus, all other information and details) was placed into the hands of programme administration (and not into programme research), this procedure allowed for a thorough separation between research and evaluation efforts on the one hand and the necessary curriculum organisation and student tracking on the other. This was communicated to students accordingly, and advantages and possible concerns regarding this procedure were openly discussed with each cohort after application, but prior to participation at pre-test.

Among the various forms of socio-communicative competence assessment – i.e. situational judgement testing, behavioural interviewing, performance-based assessment, social-cognitive assessment or personality and/or means of reflecting internal states and processes,

and portfolio assessment of work samples – situational judgement tests (SJT) seemed the preferred method for larger samples, allowing for both a high level of comparability between participants and cost- and time effective data analysis.

Micro-cultural Awareness and Change Processes.

The teaching staff in charge of programme management was vigilant in terms of cultural and group norms predominant in each cohort and maintained that attempts to implement evaluation into the curriculum require unhasty, considerate, and collaborative action and open communication. Thus, at the beginning of each new cohort, participation requirements, application forms, and information on the *PROCON* curriculum was reviewed to balance student expectations and evaluation needs. Accordingly, the name of the *PROCON* curriculum was changed from "Seminarreihe" (course of study) to "Lern- und Forschungsprojekt" (study and research project) to include this aspect. In retrospective, therefore, the level of caution in administering evaluation forms or the level of insistence on form completion can be considered as increasing with each cohort. Over time, the way in which students participated in post-module evaluation became more natural, and evaluation was generally taken for granted towards the end of the current project.

Integrative Evaluation Approach.

As described above, instruments in this evaluation included both traditional formative and summative functions. In line with the overall approach of Integrative Evaluation (Draper et al., 1996), however, the instruments were to yield utilizable information about the overall teaching and learning situation in order to inform programme management. (In this respect, it comes close to what Figl, Derntl and Motschnig-Pitrik (2005) referred to as "design-based" evaluation.) Before the development of each instrument, interviews with programme management were conducted to find out what programme management was hoping to achieve with each module, and the items derived subsequently reflect these targets. Moreover, both the results of post-module surveys and the post-tests were feed-backed and discussed with either participants or student tutors in order to reflect on and assess their validity. Inspired by a multiple action research process (e.g. McKay & Marshall, 2001; Motschnig-Pitrik, 2004; Rothman & Redlich, 2007), conclusions and outcomes of these reflections as well as changes in learning objectives were used to further develop the instruments used between cohorts. This is much in contrast to the notion held by traditional scientific researchers that any change of instruments may pose a significant threat to internal validity and may therefore warrant some explanation. While it is true that, where possible, items were held constant, each instrument

underwent some modifications. The reasons for allowing and, yet, even instigating modifications were based on the belief that these instruments were to produce results utilizable by programme management, and, consequently, had to include (a) questions based on the outcome of previous results and (b) items pertaining to subsequent changes initiated by curriculum management.

DIGEST 3.6

Three major principles guided the design and implementation of instruments.
First, for purposes of effectiveness and anonymity, computer-administered forms were used where possible to obtain information from students by use of an anonymizing individual code. For the same reason, situational judgement testing seemed an apt choice among the various forms of assessment of socio-communicative competence.
Second, micro-cultural and group norms predominant in each cohort were attended to during the implementation of evaluation into the curriculum.
And third, instruments were to fulfil both formative and summative functions.

3.7 Data Sources: Instruments and Procedures

Pre-test.

The pre-test ("Vorcheck") was introduced to the *PROCON* curriculum in 1999 when it became obvious that students possessing a stock of basic knowledge on conflict and communication as well as communicative skills (such as active listening or facilitative techniques) seemed to gain more from the curriculum than those for which the curriculum itself became the main source for acquiring basic skills. In due course, a test was developed to be employed as a self-assessment to be taken before entering the *PROCON* curriculum, offering guidelines and a "check" of the pre-requisites thought to enhance learning at later stages (Deppen, 2001).

As programme management were adamant to sustain this function of the pre-test, the leeway given for evaluative purposes could only be broadened over time. Beginning in 2002, all participants of a new cohort took a three-part *Vorcheck*, or pre-test (see Appendix A). Between all three the sections of the tests, 20-minute intervals helped students regain concentration and limited the overall burden.

The first part of the test consisted of a socio-demographic data survey, a self-efficacy rating, and the German version of the DUTCH (De Dreu, Evers, Beersma, Kluwer, & Nauta, 2001) as translated by Groß (and later published in Groß, 2004). Between 2002 and 2004, this section also included an excerpt of the Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP) (Hossiep & Paschen, 1998), a standard German personality test used for assessment in professional vocational settings. Typically this first section was completed between 30 to 40 minutes. The second part of the test incorporated knowledge tests pertaining to 1-to-1-counselling and facilitation techniques. This part was completed under time constraints, i.e. depending on the length of this section, this section had to be completed between 35- to 50-minutes.

The third part of the tests was of similar length and consisted of multimedia-based items. Students were asked to identify with the role either of a professional counsellor or a professional facilitator and write down their reaction to a number of typical situations presented by audio or video technologies. To increase reliability of the application of the test and its procedures, in this third section presentation software was used to allow for almost identical presentation and time allowances for completion in all groups.

Over time, it became clear, however, that another main function of the pre-check was to inform interested participants about what they could expect to learn in the *PROCON* curriculum. Consequently, in 2004, programme management consented to the inclusion of the situational judgement tests used at post-test. These were added to the last section of the pre-test forms, and some original multimedia items were dropped to limit the duration of the third section to less than 1 hour and 30 minutes. Moreover, students were informed accordingly, that albeit they were asked to answer to the best of their abilities, the video scenes and item demand characteristics displayed typical situations and had been added chiefly for informative purposes. It turned out that students generally welcomed this opportunity to gain more insight into the curriculum and it is thought that it also shaped their individual learning objectives.

By randomly assigning pre-test participants to one of two conditions, the 2004 cohort used to determine the parallelity between the two multimedia video-based situational judgement tests developed previously to determine participant skill level at post-testing. Pre-test participants were randomly assigned to two groups which were confronted with either the video-based situational judgement test used in 2002 or the post-test section used in 2003. Ultimately, final corrections were made in pre- and post tests and tested subsequent to the initial evaluation research phase in the 2005/06 cohort.

Online Surveys for PROCON Core Modules.

Before pre-test, it was communicated to all students interested in enrolling that while they might profit from the modules offered, in exchange for participation a certain amount of their time (i.e. approx. 45-60 minutes) would be required to fill out online survey questionnaires after each module. After completion of each module, the respective hyperlink was sent to students via electronic mail, and most participating students completed the respective questionnaires in a timely manner. Completion of the evaluation forms was made obligatory for those students intending to proceed to the subsequent module. Upon receipt of the survey data, participating students were also eligible for receiving a regular certificate of attendance.

The online surveys as detailed in Appendix B were conducted in order to better understand the module preparation phase, the "atmosphere", forms of exchange and "culture of learning" within the tutored classroom sessions. The surveys pertained to information regarding the learner's use of preparational materials, characteristics of the classroom sessions, self-reported knowledge gains in various module-specific learning objectives (as formulated by programme management), and self-efficacy ratings. In those cohorts where a preparatory

courseware was employed, additional questions addressed use and attitudes towards the courseware, as well as to the relationships between courseware and classroom training.

Moreover, various more general evaluative and (open-ended questions were used in order to inform programme management and peer-tutors of possible problems, suggestions, and circumstances valued by learners. Generally in line with the principles of action research (McKay & Marshall, 2001; Rothman & Redlich, 2007; Motschnig-Pitrik, 2004), survey results were regularly feed-backed to peer tutors in joint sessions with the author and programme management. Here, interpretations of the results and estimates as to their (social) validity were undertaken in order to plan changes of subsequent modules or cohorts.

Despite the length of the questionnaires - which admittedly put an extra burden on participants the vast majority of participants accepted that filling in the evaluation forms was part of the programme, and only few students had to be repeatedly asked to complete the survey feedbacks. Of 762 participants across cohorts and modules (counting each participant in each module and each cohort as one case), only 2 did not provide a module feedback. Some even welcomed the survey as an opportunity to reflect on their learning. Client-side java scripts were implemented to eliminate errors resulting from items overseen (or filled in wrongly) by the users. All feedbacks received were cross-checked for logical consistency across similar items and modules; also, logical errors were searched for. In one case in the *Conflict Resolution in Groups* Online Survey, a random completion of the online forms seemed highly probable (equal values across heterogeneous items in stark contrast to previous and subsequent responses by the same participant) which was thus excluded from the analyses reported below.

Courseware Log Files.

In the 2003/04 and 2004/05 blended learning cohorts, the *Conflict Resolution in Groups* module courseware recorded which task pages were accessed by users. Moreover, inputs were also stored into log-files. For purposes of evaluative analysis and courseware development, students were asked to submit these log-files to programme management on a voluntary basis. Of 80 participants in 2003, 54 (68%) feed-backed their log-files or reported that they had not used the courseware to any significant degree. In 2004, this ratio was more than 80% (36 of 43). Thus, albeit not representative, these data were available for more detailed analyses. Four indicators were calculated to determine courseware use, depending on the nature and amount of information stored in the log-files (see Table 16 below).

Table 16

Indicators of "Conflict Resolution in Groups" Courseware Use (Based on Log-file Analyses)

Indicator/Variable	2003/04	2004/05
Number of module participants	n=80	n=43
Log-file reports feed-backed	n=54 (68%)	n=36 (84%)
Page completion [L_ACPAGE]	Percentage of courseware pages accessed by user	Number of courseware pages attended to (Score) [Key: page fully completed ➔ 1 page accessed only ➔ .5]
Task completion [L_TASKC]	Percentage of input tasks completed by user	Percentage of input tasks completed by user
Task text input quantities [L_TASKL]	Sum score across all text input tasks [Key: less than 150 characters ➔ 0 ≥150 and < 500 chars. ➔ .5 ≥500 characters ➔ 1]	Sum score across all text input tasks [Key: less than 150 characters ➔ 0 ≥150 and < 500 chars. ➔ .5 ≥500 characters ➔ 1]
Sum of input lengths [L_CHARS]	Sum of text input lengths (number of characters entered) across all text input tasks	Sum of text input lengths (number of characters entered) across all text input tasks

Note. The number of participants feed-backing their log-files includes N=1 participant in 2003/04 and N=" participants in 2004/05 reporting that they had not used the courseware "to any significant degree". In these cases, all four indicators were assumed to be 0.

Post-test.

Those students interested in obtaining a distinct *PROCON* Certificate (and not just a simple certificate of attendance) were required to take the post-test ("Nachcheck"). Similar to the pre-test, post-tests also consisted of three parts. In Part I, students were asked to provide evaluative feed-back, a self-efficacy rating, and to complete the German version of the DUTCH again. In 2006, additional items proposed by the author were included in the DUTCH for test development reasons, as well as Kamentz' (2006) Learning Style Inventory (K-LSI). This section took about 30 to 40 minutes to complete.

In Part II of the post-tests, students were asked to complete content and process knowledge items (developed by the author of this study) pertaining to the theory and practices of counselling for problem-solving and conflict resolution. Time constraints limited maximum completion time to 45 minutes, albeit a majority (i.e. 80%) of the students handed in the completed section before 35 minutes had passed. The knowledge tests were simultaneously passed to three subject matter experts who were also in charge of training the *PROCON* peer tutors. Subsequent analyses of their reviews and the comments of students taking the post-test led to the exclusion of various items from the total score. Most items excluded from the Problem Solving Knowledge Subtest sum score were found to display a high degree of

ambiguity, i.e. students and experts argued that they have more than one correct answer. To increase their validity as indicators of knowledge taught in the curriculum, items were excluded from the Conflict Management Knowledge subtest where more than 40 percent of all respondents chose the "don't know" option or when less than 25 percent of all respondents chose the option originally counted as correct by test designers. (Scoring keys and definitions of total scores for the knowledge tests can be found in Appendix C, Section II. For example, a total sum score was calculated for the Problem Solving Knowledge Test [N_5] (in the same way it was calculated in the 2003/04 pre-test [V_5]. In addition, an expected sum score (V_5_{pec}, N_5_{pec}) was calculated for each participant, assuming chance-based choices with an item score based on maximum likelihood probability, i.e.:

[remaining number of multiple choice options when the "don't know" option is excluded]$^{-1}$.

To account for the items' "don't know" option, the expected score comprised only those items included also the total sum scores (V_5, N_5) the individual actually tried to give a response other than "don't know" to.

Part III of the post-test consisted of a multimedia-based situational judgement test with a duration of approximately 1 hour and 15 minutes. The situational judgement test was developed using item stems, video scenes, and multiple choice options developed by subject matter experts, namely, the authors of the *Conflict Resolution in Groups* courseware (ProKonflikt Team, 2003, 2004a). As in the pre-test, presentation software was used to increase reliability in test application. Students were asked to identify with the role of a professional consultant leading a professional project team through the various stages of Conflict Moderation (as trained in the curriculum). The participants were asked to write down their reaction to a number of typical situations presented by video technology. They were subsequently asked to rate or choose from three possible written alternatives and state the reasons for their rating or choice. While the SJT used in 2002/03 was largely based on video material later used in the courseware employed in the *Conflict Resolution in Groups* module in 2003/04, a parallel version of the first 12 items of the original situational judgement test had to be developed for post-testing of the 2003/04 and subsequent cohorts.

Traditional procedures of developing parallel tests target at producing and selecting from unidimensional item domains and test types. By contrast, however, in situational judgment tests, "multidimensionality may occur within individual items"(Clause, Mullins, Nee, Pulakos, & Schmitt, 1998, p. 195), and the target constructs are often poorly understood,

possibly rendering it "impossible … to create a test specification" (ibid.). In order to produce parallel test form, therefore, the three-step "item-cloning" procedure suggested by Clause et al. (1998) was followed. As detailed in Appendix A (C. Post-test Instruments), items in the parallel version have the same number and response set as the originals, and – albeit based on a different background story – the video-supported item stems were produced carefully in order to resemble the original as closely as possible to produce equivalent content and to replicate the individual items' multidimensionality in the alternate form. The parallel version were labelled "Mozart" and "Schütz" in line with the name of the background stories' protagonists. Two advanced students from earlier cohorts that were also peer tutors with domain-specific knowledge also monitored the production of the video sequences independently in order to produce truly twin demand characteristics.

Albeit it would have been the option of choice, it was - after due consideration - decided to spare the post-test cohorts the hassle of engaging them in additional strenuous and time-consuming research activities (without additional value to the participants). The next-best option was implemented, namely to undertake empirical investigations of parallelity were undertaken in the 2004/05 pre-test. However, the double function of the pre-test as a "check" of the necessary pre-requisites for participation and as an informative event could not be compromised. Thus, another deviation from the empirical ideal of asking participants to complete both forms to allow for matched comparisons hat to be taken into account. As an alternative, participants were randomly assigned to two groups simultaneously taking either the original or the alternate test (differing only in items 1 through 12).

The analysis of item characteristics (see Table 17 below) of the two parallel versions of the situational judgment test displays largely equivalent option selection distributions for the majority of items. For two items (MM01 and MM06), however, equivalence seemed doubtful in view of the analysis, and with significant Chi-square probabilities.

For this reason, it was decided to exclude these items from comparisons up until and including the 2004/05 post-test and to exclude them from variables representing what has been described as "pick-best" and "pick-worst-reversal" scoring algorithms (Waugh & Russell, 2005). Points were awarded for correctly identifying the most appropriate option (sum scores MM_S1R, MM_S2R) or, respectively, for identifying the least appropriate as best (sum scores MM_S1F, MM_S2F, MMS3F). (Scoring keys and definitions for the sums' algorithms are detailed in Appendix A).

Table 17

Item Equivalence of Parallel Situational Judgement Test Versions (2004 Pre-test results)

Item	MM01		MM02		MM03		MM04		MM05		MM06	
Version	M	S	M	S	M	S	M	S	M	S	M	S
Selected Option [n]												
Alt. 1	5	14	10	9	0	1	7	7	6	9	5	13
Alt. 2	7	4	1	4	18	14	4	3	6	6	15	8
Alt. 3	9	4	10	8	3	7	10	11	9	7	1	1
χ^2	6.99, p=.03		2.08, p=.35		3.08, p=.22		.19, p=.91		.83, p=.66		5.67, p=.06	

Item	MM08		MM09		MM10		MM11		MM12	
Version	M	S	M	S	M	S	M	S	M	S
Selected Option [n]										
Alt. 1	12	13	10	10	0	0	5	2	7	4
Alt. 2	5	5	2	1	11	14	6	7	4	3
Alt. 3	4	4	9	11	9	6	9	12	10	13
χ^2	.02, p=.99		.51, p=.78		.96, p=.33		1.77, p=.41		1.33, p=.52	

Variable	MM_S1R		MM_S1F		MM_S3F		MM_S2R		MM_S2F		MM_SRE	
Version	M	S	M	S	M	S	M	S	M	S	M	S
N	21	22	21	22	21	22	21	22	21	22	21	22
Mean	3.8	3.4	1.9	2.0	1.5	1.7	6.4	5.3	3.7	4.2	653.1	619.9
SD	1.6	1.5	1.2	1.2	0.9	1.0	2.0	1.8	2.2	1.7	82.3	83.7
t	.73 (p=.47)		-.39 (p=.70)		-.70 (p=.48)		1.75 (p=.09)		-.85 (p=.40)		1.31 (p=.20)	
r_{tt}	.62		.64		.42		.50		.51		.29	
L_{mvc}, χ^2	.95, 2.22		.99, 0.65		.97, 1.50		.82, 8.46**		.91, 3.81		.91, 3.86*	

Notes. "M" and "S" denote the two test versions (for "Mozart" and "Schütz" respectively, see text). Absolute frequencies (N) are reported unless noted otherwise. Equivalence is assumed where Chi-Square probability (*df=2*) values are p>.10 (two-tailed). MM_SRE is based on empirical keying at post-test (sum of option percentages across all items excluding MM_01 and MM_06). Student t-tests (*df*=41, two-tailed) were performed to test for differences of the means as all Levene-Tests remained insignificant.Reliability estimates (r_{tt}) were based on Spearman-Brown-corrected r_{12} correlation coefficients calculated by creating (21x2=42)"M"-"S"-pairs matched according to the ranked similarity of their pre-test scores. To determine ranks (i.e. the best matches), three successive criteria were used, namely (1) the lowest number of differences in option selection for items MM02…05 and 08…12, (2) the same for items MM01 through 19, and (3) the lowest absolute difference between V_MM_SRE scores. Wilks' measure of parallel test equivalence L_{mvc} was calculated by using r_{tt} estimates (cf. Lienert, 1989, p. 350) with significant χ^2 (*df*=2) = -42 (ln L_{mvc}) values being an indication of dissimilarity (* for p>.95 and ** for p>.99).

Relatively low values of Wilks' measure of parallel test equivalence (L_{mvc}) are found in MM_S2R and MM_S2F, the sum of "pick-best" "pick-worst-reversal" scoring across the whole test. This is not surprising given the fact that reliability estimates were calculated based on matched "M"-"S"-pairs with best matches chiefly determined by the lowest number of differences in option selection for only the first (parallelized) half of the test. As items MM_13 through _19 are identical in both versions of the test, the significant Chi-square values thus do not jeopardize the inference that, for the time being, the two versions of the situational judgment test can be considered parallel.

In addition, a variable MM_SRE was introduced representing the sum of all items based on empirical keying (again, excluding MM01 and MM06 for cohorts up to and including the 2004/05 cohort). Here, scores were allocated to each option according to the percentage of participants selecting the respective alternative in the post-test (a procedure termed "consensual scoring" as suggested by MacCann, Roberts, Matthews & Zeidner, 2004). However, for purposes of comparison, a common key for each item was needed to be applied in both post- and pre-test and for all versions of the test. Based on the assumption that, at post-test, participants would generally have more experience and a better understanding of the subject matter, the empirical keys were derived from post-test results. For this reason, significance in Wilks' measure of parallel pre-test equivalence (L_{mvc}) in this variable (cf. Table 17) does not disqualify the parallelity hypothesis for pre-test versions.

For versions up and until 2004/05, the respective option percentages of "Mozart" and "Schütz" versions were averaged; for the 2005/06 cohort, empirical item scores were defined as the sum of the percentage of participants rating the best-rated alternative as best plus the percentage of participants rating the least appropriate alternative as least appropriate. Here, MM_SRE represents the sum of consensual item scores across all multiple choice items.

DIGEST 3.7

Data obtained in this study come from four different sources:

(1) The pre-tests ("Vorcheck", see Appendix A) consisted of socio-demographic and background data surveys, self-efficacy ratings, knowledge tests pertaining to 1-to-1-counselling and facilitation techniques and multimedia-based situational judgement test items.

(2) In module online surveys (see Appendix B) to be completed by participants of *PROCON* core modules, students were asked to evaluate their use of preparational materials, characteristics of the classroom sessions, to report knowledge gains and provide self-efficacy ratings. Where preparatory coursewares were employed, questions addressed learner's use of and attitudes towards the courseware, as well as the links between courseware and training.

(3) *Conflict Resolution in Groups* courseware user inputs were stored in log files subsequently made available by a significant majority of participants of the 2003/04 and 2004/05 cohorts.

(4) In the post-tests ("Nachcheck", see Appendix C), participants were asked to provide evaluative feed-back and self-efficacy ratings as well as information on a number of personal background variables. and to complete knowledge tests pertaining to the theory and practices of counselling for problem-solving and conflict resolution. The post-tests also comprised one of two parallel versions of a video-based situational judgement test of Conflict Moderation skills.

3.8 Data Analysis, Statistical Procedures, and Effect Size Calculation

Principles of Data Analysis.

Above and in the following, both qualitative methods as well as quantitative statistical analyses are used to evaluate the information and data collected over time. Naturally, the methods used should allow for drawing valid conclusions about the guiding assumptions for data analysis summarized above (see section 2.8 above).

This may be the sole principle guiding much of the research undertaken in the fields of psychology and education. By contrast, the author of this study maintains, however, that additional strategic objectives have to be taken into consideration. Evaluative studies traditionally aim at informing programme management or stakeholders. One must therefore also be concerned with both the choice of statistical methods and the way the analyses are conducted. (Ideally, in an evaluative context, the presentation of the results would also be predominantly guided by this principle. Conversely, given the type of this document, the author is largely bound to comply with both academic standards and the current requirements set by reviewing and authorizing bodies for the academic degree pursued.)

In practice, there may be mathematical analyses available that render possible complex analyses of research questions. However, they may tend to require a considerable deal of explanation in order to be comprehensible to the stakeholders. This should not be taken to mean that evaluative researchers should not perform higher-order statistical procedures. However, the amount of clarifying explanation needed may sometimes outweigh the possible informational gain attained by a "scholarly apt" choice of a mathematical or statistical model. Thus, it is argued here that procedures for data analysis should be selected on a case-by-case basis.

Moreover, much of all evaluative research undertaken in the fields of psychology and education, the quantitative data and information produced has chiefly ordinal justification (Cliff, 1996). The same holds true for most of the data produced in this evaluative study as most Online Survey items are based on (discrete) Likert scales. However this may also apply for aggregate (sum) scores traditionally believed to approximate continuous data and distributional models dependent thereon.

For various reasons, the main characteristic of S. Stevens' (1946) conception of an interval level of measurement, namely, additivity, cannot be automatically assumed. As in

many evaluative studies, the vast majority of measures used had to be created from scratch without being able to control their psychometric properties, and without reference to an underlying statistical model. Moreover, the individual items of composite measures (such as knowledge or situational judgement tests) typically measure different aspects of the underlying ability. And, in addition, there currently is insufficient information as to the items' predictive or construct validity; too little is known about their relative worth or significance in terms of the construct intended to measure to reliably estimate their individual relative weights. In other words, while the sum scores may display certain distributional characteristics, they may only give an ordinal estimate as to the competencies measured. They may "convey little or no information about what empirical relations the numbers represent", as Long (2005, p. 330) has pointed out.

Despite these concerns, many statistical procedures can be reliably computed as long as minimal distributional properties (such as, normality) are given. However, when using higher-level of measurement statistical procedures for lower-level of measurement data, the results of statistical tests may be received with suspicion, or they altogether be meaningless due to the "mismatch between the information represented by the numbers and the nature of the model parameter on which the test is based" (Long, 2005, p. 329).

To avoid these justificational and interpretational problems, a more conservative approach to statistical data analysis was used in the following:

(1) All data reported below were screened for outliers, and tested for distributional properties (such as normality).

(2) Where interval level of measurements could not be sufficiently be justified, ordinal level of measurements were assumed, and statistical procedures were adapted accordingly.

(3) Where parametric tests are traditionally considered suitable for the level of measurement of the data given, these were used when tests of violations of their major mathematical assumptions (e.g. homogeneity of variance) yielded negative results. When violations distributional or major mathematical assumptions could not be ruled out, the results of distribution-free (non-parametric) statistical tests are reported in addition.

(4) Where salient violations of parametric tests could not be ruled out, non-parametric alternatives were used in place of parametric tests. (In this case, the results of traditionally used parametric statistical procedures are sometimes also presented.)

(5) Current reporting standards (APA, 2001) call for the inclusion of indices reflecting the magnitude of effects or the strength of relationships. In line with the comprehensibility principle outlined above, it was decided to predominantly draw on one standard effect size throughout in the following, and regardless of the specific type of inference one may be concerned with in a specific case (i.e. relationship/association or difference). In addition, a single standard effect size estimate may also facilitate the drawing of inferences based on different examinational procedures within this study. (The same principle is followed in meta-analyses for the reason of integrating the findings from different sources, but it can also be applied here.) Possibly, the class of effect sizes indices most easily interpretable by stakeholders and programme management across the various types of type of inference are those concerned with between-group magnitude of difference. The most prominent indices of this class possibly is Cohens' *d* (or its respective sample analogues such as *g;* e.g. Cohen, 1988) which represents the difference of means expressed in units of standard deviation. Cohen's *d* is therefore used as a standard placeholder for effect size estimates reported below. In the following, a description is given on how this "standard effect size" estimates reported in the sections below is derived, and an index is affixed to *d* to indicate on which formula its calculation is based. If not noted otherwise, all estimates will be based on one-tailed test statistics with two-tailed confidence-intervals (CI) based on converted CI boundaries of the original test statistic.

Parametric Effect Size Estimates.

(1) Where independent groups are compared, and the hypotheses may be upheld that mathematical assumptions of parametric statistical tests are not saliently violated, or where parametric statistical tests have been shown to be robust to minor violations of its mathematical assumptions, Cohen's *d* is calculated as d_t=t $((1/n_j)+(1/n_k))^{1/2}$, where t reflects the mean difference between the groups j and k, divided by the pooled standard error of difference for unbalanced designs and n_j and n_k the number of participants in each group. Confidence intervals (CI) are calculated based on $CI = d_t \pm z_{\alpha/2}\sqrt{\left(\frac{1}{n_j}+\frac{1}{n_k}\right)+\frac{d^2}{2n_jn_k}}$, where $z_{\alpha/2}$ denotes the critical value from the normal distribution at the given two-tailed α-level cumulative density. For example, a 95% CI would use $z_{97.5\%}\approx 1.95995$.

(2) Typically, within-cohort comparisons are correlated designs in that samples are drawn repeatedly from the same group of participants. Due to the correlation between the dependent samples, however, the estimate of *d* must be corrected to avoid overestimation.

Here, the procedure put forward by Dunlop, Cortina, Vaslow, and Burke (1996) is used and the effect size is calculated as $d_{t_c} = t_c\sqrt{2(1-r)/n}$, where t_c is the value of the t-test for two dependent samples, r is the correlation between measures, and n is the sample size. For this effect size estimate, confidence levels are calculated using a variance estimation equation derived by Becker (1988, p. 261; cit. by Dunlop at al., 1996, p. 172) which also accounts for the correlation between measures: $CI = d_{t_c} \pm z_{\alpha/2}\sqrt{\frac{2(1-r)}{n} + \frac{d^2}{2n-2}}$.

(3) Cohen's *d* values were calculated as $d_r=(4r^2/(1-r^2))^{1/2}$ where the test statistics are based on the class of parametric Pearson correlation coefficients (r, r_{pb}, r_{tet}, r_b). The confidence intervals for these are typically based on Fisher's-*Z*-transformations $Z_r=\ln(1+r/1-r)/2$ and $r=(e^{2Zr}-1)/(e^{2Zr}+1)$ and with a $CI_{Z_r}=Z_r\pm z_{\alpha/2}(1/N-3)^{1/2}$.

Effect Size Estimates based on Non-Parametric Statistics.

Common parametric effect size estimates are bound to be sensitive to violations of their underlying assumptions, such as differences in distribution, heterogeneity of variance, sample size differences, or small sample biases (2001; cit. in Leech & Onwuegbuzie, 2002). Thus, effect size estimators for non-parametric between-group and repeated-measures comparisons have been developed which also allow for conversions to the parametric Cohen's *d* index. The estimators used in this study are outlined below.

(4) Cohen's *d* may be calculated from Spearman rank-order (rho) coefficients r_s as $d_{r_s}=2r_s(1-r_s^2)^{-1/2}$. The CI values for r_s are principally based on the same Fisher-*Z*-transformation used for Pearson-class *r*s, with a correction factor suggested by Zar (1999; cit. in Sheskin, 2007, p. 1364): $CI_{Z_{r_s}}=Z_{r_s}\pm z_{\alpha/2}(1/1.060/N-3)^{1/2}$.

(5) For ordinal statistics, a class of non-parametric approaches is based on ratio indices (Grissom, 1994; Grissom & Kim, 2001; Kraemer & Andrews, 1982; Romano, Kromrey, Coraggio, & Skowronek, 2006), one example of which is what McGraw and Wong (1992) have termed the *common language effect size*. Here, the probability that one sampled member of the (treated) population will have a dependent variable score that differs in a predetermined fashion, i.e. is higher ($Pr(x_1>x_2)$) or lower ($Pr(x_1>x_2)$), than that of another population treated differently. This general approach has later been called the *probability of superiority* (PS), and has been found robust to non-normality (Grissom & Kim, 2001). For two independent samples, the probability value is based on all possible between-subject comparisons. The number of

instances in which the score of the i^{th} participant in sample 1 (x_{i1}) is greater than the one of the jth participant of sample 2 is divided by the total number of comparisons undertaken.

Dominance, a delta statistic originally suggested by Cliff (1993, 1996b) also takes into account the number of reverse occurrences. This may be of special importance where participants develop contrary to the intentions of those intervening. In its general form for independent samples, it is calculated by comparing the scores of all n_i cases in sample i with the scores of all $1…n_k$ cases in sample k: $\delta = (\#(x_i>x_k) - \#(x_j<x_k)) / (n_j n_k) = d_{ik}/(n_j n_k)$. In this so-called dominance matrix, all entries $d_{ij}=sign(x_i-x_k)$. In can also be derived from Mann-Whitney's U statistic with $\delta =(2U/ n_i n_k)-1$ (Cliff, 1996, p. 495).

This ordinal delta statistic ranges from $\delta = +1.0$ (in case the scores of sample 1 exceed all sample 2 scores) to $\delta = -1.0$ (for the reverse). It is also easily interpretable as the 'net' percentage of scores being higher in one group than in the other. As Hogarty and Kromrey (2003) point out, Cliff's delta appears to be most robust. It may be especially apt for "small and moderate samples in terms of Type I error rate, power, and coverage of the CI … " as well as in conditions of "nonnormality, heterogeneity, and unequal sample sizes." (Feng & Cliff, 2004, p. 331).

For independent samples, the asymmetric CI formula put forward by Feng and Cliff (2004) was used to calculate confidence intervals for both δ and d_δ with

$$CI_\delta = \frac{\delta - \delta^3 \pm z_{\alpha/2} s_\delta \sqrt{1 - 2\delta^2 + \delta^4 + z_{\alpha/2}^2 s_\delta^2}}{1 - \delta^2 + z_{\alpha/2}^2 s_\delta^2}, \text{ where } s_\delta^2 = \frac{(n_i - 1)s_{d_{i.}}^2 + (n_j - 1)s_{d_{.j}}^2 + s_{d_{ij}}^2}{n_i n_j}$$

based on the dominance matrix' variances of rows and columns averages.

Cliff (1996a, 1996b, 1996c) and Long, Feng, and Cliff (2003) also put forward the counterpart for paired data. When all x_i (post-test) scores of all n participants with all x_k (pre-test) scores are compared ($\Sigma\Sigma\ d_{ik}/n^2$), information about individual (within-pair) changes ($\delta_w=\Sigma d_{ii}/n$) is mixed with information about between-pair changes, i.e. the change of the group as a whole ($\delta_b = (n(n-1))^{-1} \sum_{i \neq k} \sum d_{ik}$).

Ordinary CI formulae are used in for paired comparisons as suggested by Long, Feng, and Cliff (2003) with

$$CI_{\delta_w} = \delta_w \pm z_{\alpha/2} \frac{1}{\sqrt{n}} \sqrt{\frac{\sum (d_{ii} - \delta_w)^2}{(n-1)}} \text{ and } CI_{\delta_b} = \delta_b \pm z_{\alpha/2} s_{d_b}.$$

For multiple comparisons, a step-down Bonferroni procedure was used to control the family-wise α-level as suggested by Cliff (1996c, p. 168).

Romano et al. (2006) suggest that both δ and Cohen's *d* may be interpreted in terms of non-overlap between normal distributions, thus allowing for δ values be transformed and expressed in *d* terms. For this purpose, the following procedure is used: First, inversely, the mean difference z point is sought where the cumulative area under the normal distribution equals -1/(|δ|-2). Then, respective z value is doubled to find an equivalent to Cohen's *d* which will be referred to in the following as d_δ. For example, a delta value of δ=-.15 would be equivalent to d_δ=.204, where δ=.329 *d* approximates .5, and when δ=.50, d_δ=.861. For dependent sample, Cohen's *d* is calculated based on the δ_b statistic.

As delta statistics were not readily available by the version of SPSS used, all calculations were performed using SPSS macro extensions developed by the author along the formulae, examples, and procedures detailed in Long, Feng, and Cliff (1998, 1999, 2005; Long et al., 2003). It may be found in Appendix E.

Interpretations of Effect Size.

To allow for better comparisons below, the following Table 18 indicates how Cohen's d may be interpreted in comparison to the aforementioned other effect size estimators. Cohen himself tentatively stated that "there is a certain risk in inherent in offering conventional operational definitions for those terms for use in power analysis in as diverse a field of inquiry as behavioral science" (J. Cohen, 1988, p. 25); however, he suggested that d values greater than .2 but not more than .5 may be termed "small", values greater than .5 but not more than .8 "medium", and values greater than .8 be called "large" (1988, pp. 24-27). The following Table 18 provides an overview for the effect size estimators outlined above.

Table 18

Appraisal of Effect Size Estimators

ohen (1988)	Cohen's *d*	Percentile Normal Distrib.	p (one-tailed)	δ Cliff (1993)	Pr(Y>X) Grissom (1994)	r
	0.0	50.0%	.500	.0000	.5000	0.000
"small	0.2	57.9%	.421	.1476	.5562	0.100
effect size"	0.4	65.5%	.345	.2737	.6114	0.196
"medium	0.5	69.2%	.308	.3297	.6382	0.243
effect size"	0.6	72.6%	.274	.3816	.6643	0.287
"large	0.8	78.8%	.212	.4743	.7142	0.371
effect size"	1.0	84.1%	.159	.5538	.7602	0.447
	1.2	88.5%	.115	.6621	.8019	0.482
	1.4	91.9%	.081	.6808	.8389	0.573
	1.6	94.5%	.055	.7312	.8711	0.625
	1.8	96.4%	.036	.7744	.8985	0.669
	2.0	97.7%	.023	.8114	.9214	0.707
	2.33	99.0%	.010	.8610	.9503	0.759
	3.10	99,9%	.001	.9355	.9858	0.840

DIGEST 3.8

It is argued that both the choice of statistical methods and the way the analyses are conducted should not solely be guided by the traditional objectives of research. The possible informational gains attained by a "scholarly apt" choice of a mathematical or statistical model may sometimes be outweighed by the amount of clarifying explanation needed in order to be comprehensible to stakeholders. An ordinal level of measurements is assumed for the majority of data obtained in this study to avoid interpretational and justificational problems, choosing and adapting statistical procedures accordingly where necessary. In line with the comprehensibility principle outlined above, it was decided to predominantly use Cohen's d as a standard placeholder for effect size estimates reported.

4. Results

It was concluded above that little is known about the question, why and in what ways technology enhance can training programs for soft skills. Both theoretical considerations as well as empirical findings led to the assumption that what has been described as Computer-Supported Social Learning (CSSL) may render fruitful effects.

Eight summative assumptions were detailed above which, in short, further define the notion that where CSSL courseware is used as a supplementary means of preparation for extant soft skills in-class trainings, augmented increases and higher resulting levels of socio-communicative competence may be effected.

The rationale and background of this study was detailed in the foregoing. Conflict-resolution-related CSSL courseware was introduced into a university-based, skills-oriented curriculum, and comprehensive descriptions were given as to evaluation design, the study participants, the instruments used, and principles of data analysis.

In the following, empirical findings and the results of statistical analyses addressing the eight assumptions are reported.

4.1 Indicators of Learning Gains

The *PROCON* curriculum was introduced primarily to allow students to develop key socio-communicative competencies in the fields of facilitation, problem solving and conflict resolution. One main question thus pertains to the questions of what and how much do students actually learn in the various modules? The evaluation system presented above aims at providing answers to these questions. Various indicators were introduced to learn more about learning gains, the characteristics and nature of competencies that learners actually developed and to provide quantified indicators of progress and learning gains. These are detailed in the following.

Self-Reported Learning Gains.

In each post-Module Online Survey, participants were asked for their estimate of learning gains in pre-defined areas ([F09], [P09], [M09], [K09]; see Appendix B for details). These items were introduced to directly and indirectly address likely and possible learning objectives. They were developed based on an analysis of the literature and courseware contents as well as a series of interviews with programme management and teaching staff. Over time, the item set was adapted and supplemented by items to reflect changes in the curriculum or workshop adaptations recommended by previous students and peer-tutors. The participants were asked to rate their respective gain on a 6-point Likert scale with "1" denoted as "very high gain" and "6" as "very low gain" to explore the modules' actual effects in terms of learning content and achieved learning objectives.

The results are summarized in Figures 18 and 19 below. Across all items and cohorts, the mean values for self-rated knowledge gains range between M_{min}=1.95 and M_{max}=3.47 meaning that participants generally feel that they have had a benefit from taking part in a module. Information contained in Figures 18 and 19 may also be treated as a formative to guide possible changes of the curriculum. For this purpose, major findings for the core modules will be outlined in the following.

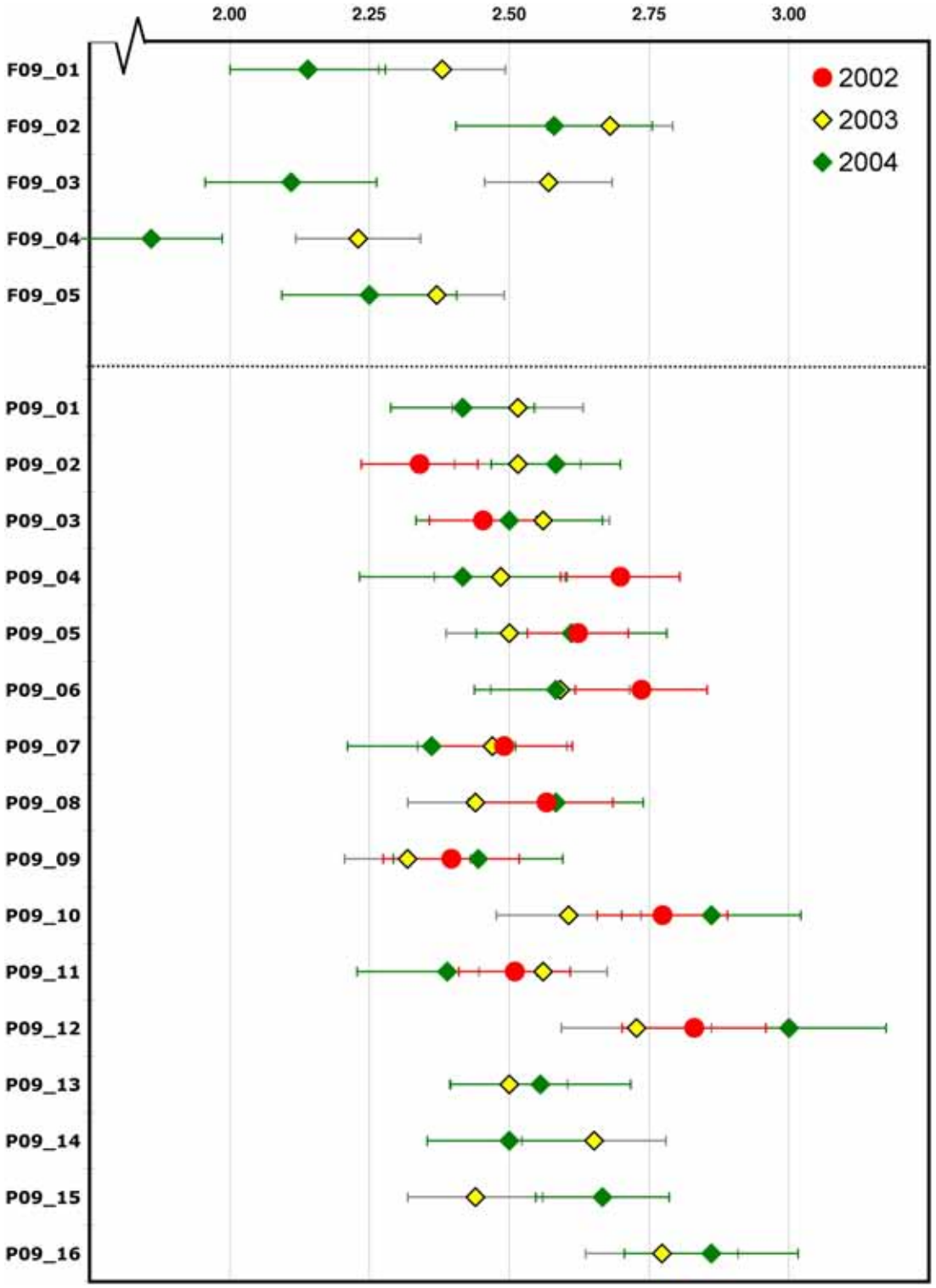


Figure 18. Self-rated skills and knowledge gains in *Facilitation / Problem Solving* Module Online Surveys (see Appendix B for details).

Notes. Points represent the mean of items rated on Likert scale (1 -"very high gain" … 6 – "very low gain") for sample base cohorts 2002/03, 2003/04, and 2004/05. Horizontal lines depict standard errors of the means. Note that a curtailed metric grid was used for purposes of clarity; distances may appear aggrandised.

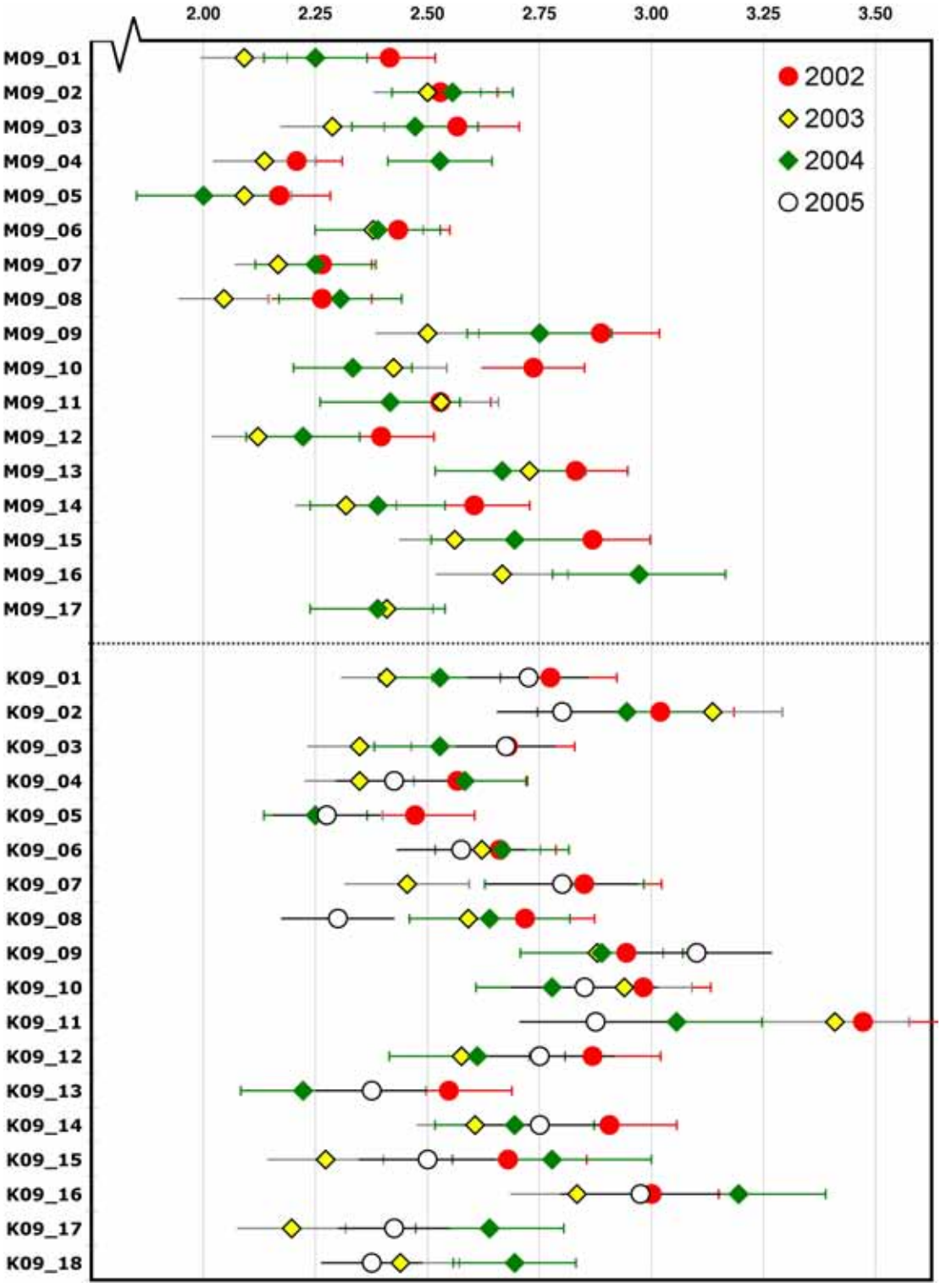


Figure 19. Self-rated skills and knowledge gains in *Mediation Skills* and *Conflict Resolution in Groups* Module Online Surveys (see Appendix B for details).
Notes. Points represent the mean of items rated on Likert scale (1 -"very high gain" … 6 – "very low gain") for sample base cohorts 2002/03, 2003/04, and 2004/05. Horizontal lines depict standard errors of the means. Note that a curtailed metric grid was used for purposes of clarity; distances may appear aggrandised.

On the whole, the greatest benefits of the *Problem Solving* modules seem to lie in "Root-cause hypotheses generation and testing" [P09_07] and "Gathering, classifying and

weighting criteria for solution valuation" [P09_09]. On the other hand, determining yet which of the problem solving interventions taught are indicated in a case [P09_12] seems to be most problematic even after participation. The same holds true for "Conducting a thorough Potential Problem/opportunity Analysis" [P09_10].

As for the *Mediation Skills* modules, the highest average gains across cohorts were reported for "Structuring own actions and the process of mediation" [M09_05], " Deepening understanding of backgrounds through bilateral active listening" [M09_07] and "Promoting and securing fairness for both parties" [M09_08]. By contrast, items with the lowest relative mean values across cohorts were "Delineate and phrase conflicting issues" [M09_09] and "Concise and fluent visualisation of issues/subject matter" [M09_10].

In the *Conflict Resolution in Groups* modules, maximum average skill increases were reported for the "process structuring" items [K09_05; K09_13], "Clarifying the mediator's role and obtaining co-operation agreements with the parties involved" [K09_04]. Participants stated that they obtained least average gains in the areas of "Promoting implementation of solutions arrived at" [K09_11], "Instigating other's inclination to compromise" [K09_16], "Conducting a complete preliminary meeting with the team leader" [K09_02], and, as above, "Delineate and phrase conflicting issues" [K09_09].

It was found repeatedly in meetings and discussions that these outcomes are generally in line with expectations voiced by peer-tutors and module supervisors; thus it may be argued that the self-reported learning gains enjoy a sound level of content validity and do reflect changes as perceived by close observers of the learning processes.

Changes in Self-Efficacy Levels.

In Pre- and Post-tests as well as in all Module Online Surveys, participants were repeatedly asked to evaluate and rate the scope of their knowledge and skills in various areas. In Figures 20a through 20d indicate that major changes in self-efficacy levels (and thus, learning gains) occurred just when students had taken part in a module set up for the very purpose of skill development in that respective area.

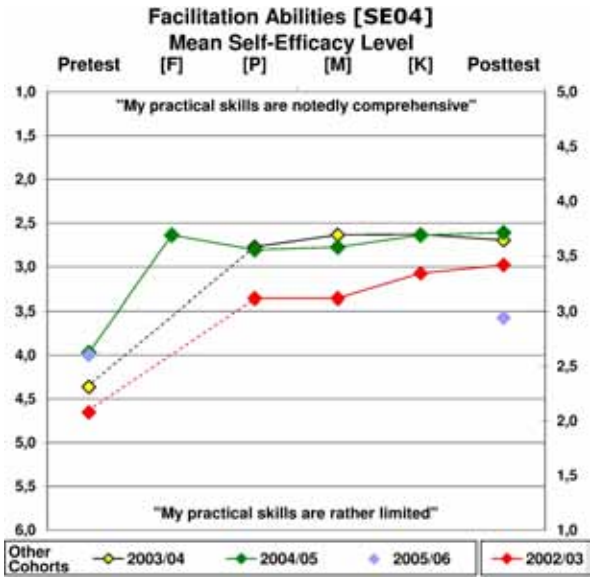


Figure 20a.

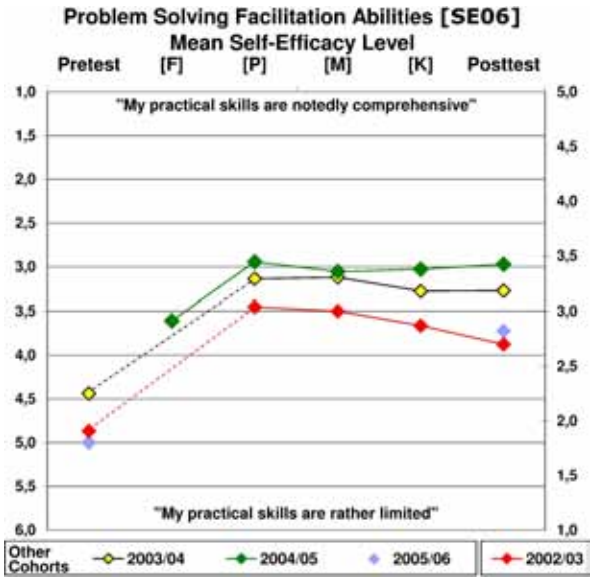


Figure 20b.

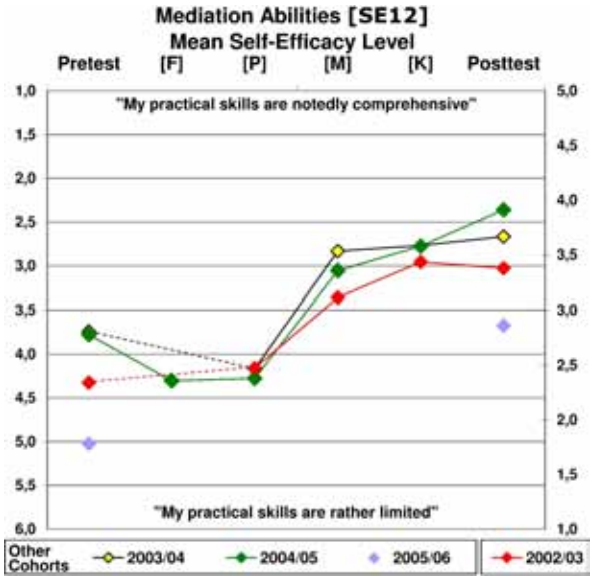


Figure 20c.

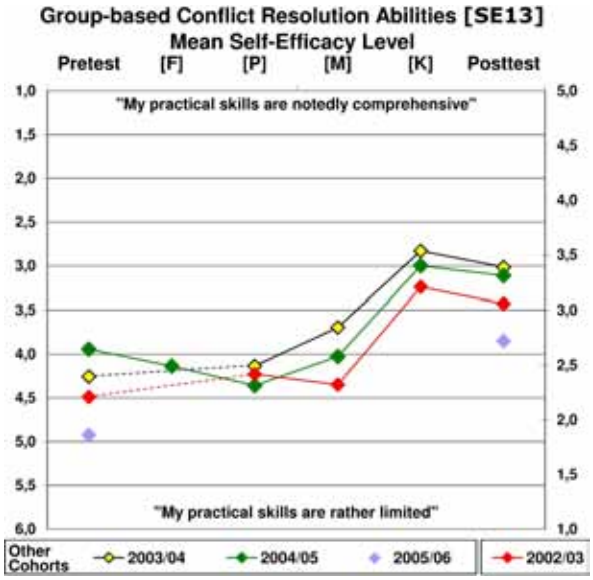


Figure 20d.

Albeit mean values are reported in Figures 20a to 20d, the underlying distributions are naturally not continuous but based on discrete Likert scale data at ordinal level. Here, equal interspaces cannot be automatically assumed. Thus, attempting to treat individual items and using statistical procedures intended for interval or ratio data is questionable, also since most of these statistics require samples drawn from normally distributed populations. Testing for normality on Likert item level, however, is arduous and again very problematic as due to the relatively small subsamples the expected frequencies in are either extremely small (which leads to salient violations of the Chi-Square goodness-of-fit tests) or re-grouping leads to arbitrary categories and a loss of power of alternative procedures such as, for example, the Kolmogorov-Smirnov goodness-of-fit test for discrete ordinal data (Pettitt & Stephens, 1977). To by-pass

these potential problems, paired comparisons were tested for median differences using the Wilcoxon matched-pairs signed-ranks test and Cliff's (1996, Long et al., 2003) delta dominance statistic for ordinal correlated data.

As suggested above (see Figures 20a-d), some gains self-efficacy in self-efficacy may be corresponding to subject matter taught across certain time spans in the *PROCON* curriculum or in specific modules thereof. Figure 21 denotes these hypothesized relationships.

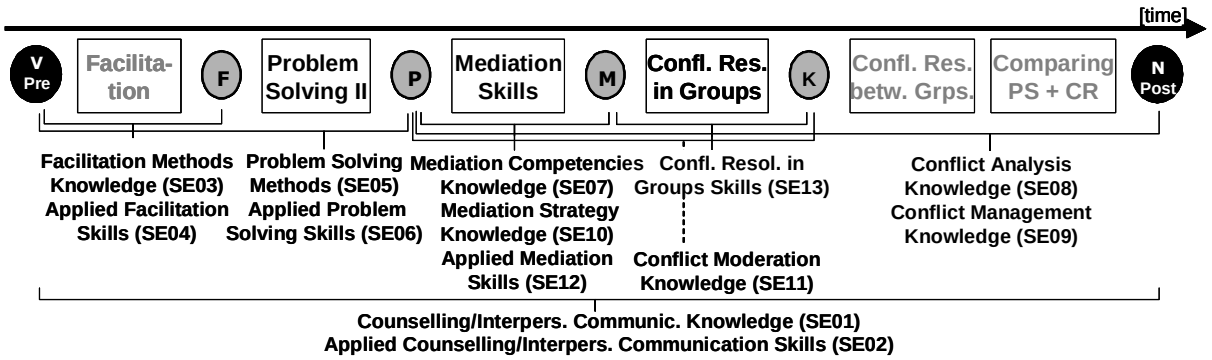


Figure 21. Hypothesized links between *PROCON* modules and gains in self-efficacy as reported.

Theoretically, to give some first background as to the validity of a claim that gains self-efficacy may, indeed, be linked to certain modules or sequences of modules, a paired comparison of interest (PCOI) should display a significant increase in self-efficacy. Preferably, this should apply to all cohorts tested. Moreover, while non-significant increases only indicate that differences possibly existing in reality have not been detected, non-significant differences between the complementary paired comparisons of subsequent surveys (e.g. V-F/P, F-P, P-M, M-K, K-N) it may also indicate that differences may not be substantially influenced by these modules. Preferably, therefore, where applicable, no significant gains should be found with the exclusive exception of the paired comparison of interest (PCOI), based on a Bonferroni-Dunn-corrected $p<.05$ family-wise Type I error rate. The results of this analysis are reported in Table 19.

Table 19

Sample Base Self Efficacy Gains in PROCON *Modules with Corresponding Subject Matter*

SE-Item/Content Desciption	PCOI[d]	2002/03[a]			2003/04[b]			2004/05[c]		
		ΔM[e]	z[f]	other n.s.[g]	ΔM[e]	z	Other n.s.[h]	ΔM[e]	z	other n.s.[h]
01 Counselling/Interpers. Knowledge	V-N	.9	5.00***	-	1.6	-6.27***	-	1.0	-3.61***	-
02 Applied Counselling/Interpers. Skills	V-N	1.0	4.77***	-	1.2	-5.28***	-	.8	-3.44***	-
03 Facilitation Methods Knowledge	V-F[h]	.8	4.57***	Yes	1.6	-6.25***	Yes	1.6	-4.44***	Yes
04 Applied Facilitation Skills	V-F[h]	1.0	5.50***	Yes	1.6	-6.62***	Yes	1.3	-4.60***	Yes
05 Problem Solving Methods Knowledge	V-P[i]	1.2	5.63***	Yes	1.5	-6.35***	Yes	.4	-2.00*	Yes
06 Applied Problem Solving Skills	V-P[i]	1.1	5.51***	Yes	1.3	-5.84***	Yes	.7	-3.10**	Yes
07 Mediation Competencies Knowledge	P-M[j]	.9	3.95***	Yes	1.4	-6.09***	Yes	1.6	-4.60*	No[k]
08 Conflict Analysis Knowledge	P-N	.5	2.92**	-	.5	-3.10**	-	1.4	-4.24***	-
09 Conflict Management Knowledge	P-N	.9	4.26***	-	1.3	-5.99***	-	1.4	-4.17***	-
10 Mediation Strategy Knowledge	P-M[j]	1.5	5.30***	Yes	1.8	-6.20***	Yes	1.7	-4.64***	Yes
11 Conflict Moderation Knowledge	P-K[j]	1.3	5.54***	-	1.8	-6.54***	-	2.0	-4.82***	-
12 Mediation Skills	P-M	.6	3.63***	No[l]	1.3	-5.73***	Yes	1.2	-4.16***	Yes
13 Conflict Resolution in Groups Skills	M-K	.9	4.39***	Yes	.9	-4.73***	Yes	1.0	-3.77***	Yes

* p<.05. ** p<.01. *** p<0.001.

[a]n=53. [b]n=66. [c]n=35. [d]Paired comparison of interest, i.e. stating which of all possible paired comparisons should show significant gains in self efficacy if gains correspond to the subject matter of certain *PROCON* modules (with V-Pre-test, F-Facilitation module online survey, P-Problem Solving module online survey, M-Mediation Skills module online survey, K-Conflict Resolution in Groups module online survey, N-Post-test).

[e]Difference of means in paired comparison of interest (calculated as $\Delta M = M_{post} - M_{pre}$ for 2002/03 and $\Delta M = M_{pre} - M_{post}$ for other cohorts). Positive values denote an average increase in self-confidence.

[f]Normal approximation of Wilcoxon matched-pairs signed-ranks T Test statistic.

[g]Indicates if - with the exclusive exception of the paired comparison of interest listed in the PCOI heading - all other paired comparisons of subsequent surveys (V-F/P, F-P, P-M, M-K, K-N) were not found to be significant (based on a Bonferroni-Dunn-corrected p<.05 family-wise Type I error rate). Not applicable for pairs with intermediate surveys (V-N, P-N).

[h]For cohorts without Facilitation modules (F), paired comparisons were conducted with postliminary self-efficacy ratings of the subsequent Problem Solving online survey (V-P).

[i]Paired comparisons were conducted with preliminary self-efficacy ratings of the preceding Facilitation online survey (F-P) for the 2004/05 cohort.

[j]Not included in pre-tests (V) and post-tests (N).

[k]Further significant increase in/after *Conflict Resolution in Groups* Module (M-K, r=.285, p=.016).

[l]Further significant increase in/after *Conflict Resolution in Groups* Module (M-K, r=.258, p=.019).

As Table 19 indicates, significant increases in self-efficacy levels predominantly occur at certain times within the curriculum, and with few exceptions, across all cohorts surveyed. It may thus be argued that - although these findings are not based on truly experimental conditions – it may not be unreasonable to believe that the changes in self-efficacy may be related to participant change in individual modules or blocks of modules.

The estimated magnitudes of change are shown in Figures 21a through 21d. According to the benchmark labels suggested by J. Cohen (1988, see Chapter 3.8 above) these may be regarded as “medium” to “large” effects.

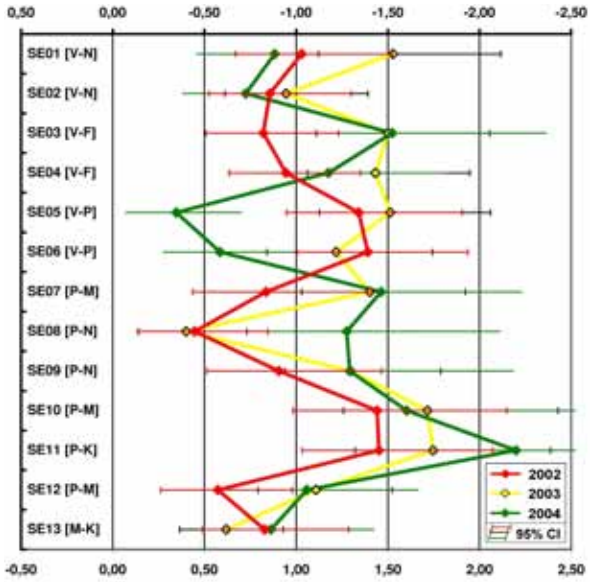


Figure 21a.

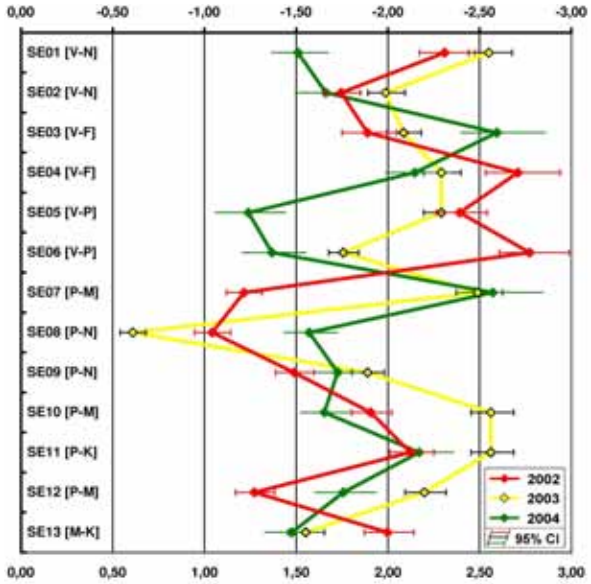


Figure 21b.

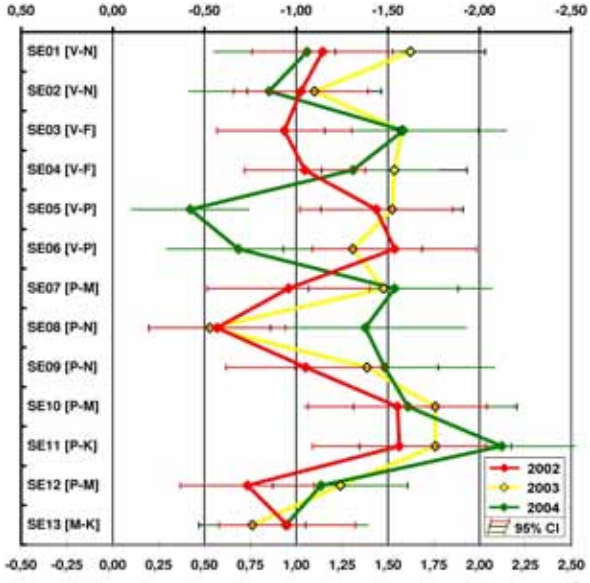


Figure 21c.

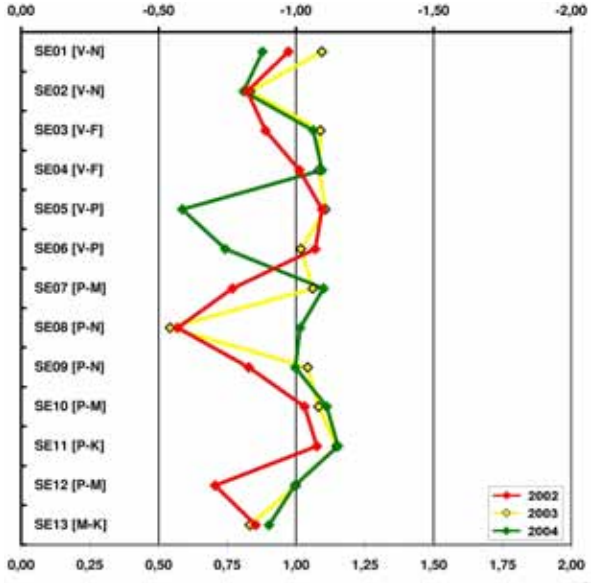


Figure 21d.

A more detailed analysis of the effect sizes revealed that, of the three non-parametric statistics, Cliff's delta (δ_b) yields a Cohen's *d* estimation (Figure 21a) most closely related to the one based on the t-test for dependent samples (Figure 21c). This finding underscores the attractive characteristics of the delta statistic conveyed by current statistics research (Feng, 2007). For purposes of comparison, Probabilities of Superiority (PS) were calculated in accordance with Grissom and Kim (2005, pp- 114-115) as cited and exemplified by Sheskin (2007, p. 801) and then re-transformed to effect size estimates by use of procedures detailed in Grissom (1994, p. 315). In contrast to the delta statistic, the PS statistic (Figure 21b) tends to over-estimate the group effects. This is probably due to the concentration of within-subject

change only, by not accounting for scores contrary to the ones analysed, and despite the fact that the PS in dependent samples does take ties, i.e. zero differences, into account.

As Figures 21a-d suggest, for some *PROCON* Modules and cohorts, the magnitude of increases in self-efficacy may differ significantly between cohorts. First, increases (as measured in Cliff's ordinal between-subject δ_b) in item SE03 (knowledge about methods and procedures used in facilitation) are significantly lower in cohort 2002/03 when compared to both 2003/04 ($z_{\Delta\delta}=\delta_{b1}-\delta_{b2}/(s_{\delta b1}{}^2+s_{\delta b2}{}^2)^{-1/2}=2.340$, p=.010) and 2004/05 ($z_{\Delta\delta}=2.108$, p=.018). (This may be explained with the introduction of an additional (blended-learning) module in 2003/04 aimed at helping participants develop knowledge about methods and procedures used in facilitation.)

Second, the self-rated extent to which a participant has acquired knowledge about competencies and skills needed for mediation (SE07) in the *PROCON* modules *Mediation Skills* [M] and *Conflict Resolution in Groups* [K] has increased to a significantly higher degree in both blended learning cohorts 2003/04 ($z_{\Delta\delta}=1.660$, p=.048) and 2004/05 ($z_{\Delta\delta}=1.663$, p=.048). It will be ultimately a matter of further investigation whether these differences may be explained by the introduction of computer-supported learning.

Third, the same applies to the practical mediation abilities self-efficacy rating (SE12) pre and post the *PROCON Mediation Skills* [M] module with $z_{\Delta\delta}=2.040$ (p=.020) for 2002/2003 and ($z_{\Delta\delta}=1.639$, p=.050) for the 2002/2004 comparison.

And, finally, the 2004/05 cohort shows significantly higher self-efficacy gains in item SE 08 (self-rated knowledge about antecedents and sources of inter-personal conflict) when compared to both 2002/03 ($z_{\Delta\delta}=2.529$, p=.006) and 2003/04 ($z_{\Delta\delta}=2.902$, p=.002).

Knowledge and SJT Test Scores.

Another main group of quantified indicators of learning are post-test knowledge and situational judgement test scores. Test performance scores should be significantly higher than what could have been achieved by chance or what had actually been achieved in the pre-test. The respective results are presented in Table 20 below.

As far as pre-post situational judgement test comparisons are concerned, medium to large improvements in identifying the most appropriate answers ("pick-best"; cf. Waugh & Russell, 2005) were found. The same holds true for the item score MM_SRE based on empirical consensual keying (MacCann, Roberts, Matthews & Zeidner, 2004) where scores were allocated to each option according to the percentage of participants selecting the

respective alternative in the post-test. This applies to both the 2004/05 cohort (ordinal δ_b=.797, p<.0000 with Cohen's d_{δ_b}=1.917; interval t_c=8.692, *df*=35, p<.0001, Cohen's *d*=1.856 with a 95% CI of 1.252 to 2.459) and the 2005/06 cohort (ordinal δ_b=.658, p<.0001 with Cohen's d_{δ_b}=1.319; interval t_c=8.213, *df*=39, p<.0001, Cohen's *d*=1.336 with a 95% CI of .901 to 1.771).

No significant changes were observed for incorrect answers sum scores MM_S1F and MM_S2F. However, significant declines were found in cohorts 2004/05 (Wilcoxon T(36)=725, Z_T= -3.963, p<.0001; δ_b= -.511, p<.0001 with Cohen's d_{δ_b}= -.889) and 2005/06 (Wilcoxon T(40)=75, Z_T= -3.762, p<.0001; δ_b=-.464, p<.0001 with Cohen's d_{δ_b}= -.777) for the 'salient errors' sum score MM_S3F. This score denotes the number of occurrences in which the least appropriate alternative was incorrectly taken for the best ("pick-worst-reversal"; cf. Waugh & Russell, 2005).

Table 20

Knowledge and Situational Judgement Tests: Differences between Pre-test and Post-test

Test	Cohort	N	Variable	Pre [V][a]	Post [N][a]	T[b]	d_{δ}[c]	$CI_{95\%}$ d_{δ}	
								lower	upper
Multimedia-Based Situational Judgement Test (Post-test Part III, see Appendix C)	2004/05	36	MM_S1R	3.4 (1.4)	6.3 (1.1)	0 ***	2.520	1.935	3.739
			MM_S1F	2.0 (1.2)	2.2 (1.0)	165	-.121	-.460	.165
			MM_S2R	5.6 (1.9)	8.8 (1.9)	131 **	1.870	1.321	2.864
			MM_S2F	4.2 (1.9)	4.2 (1.5)	261	.006	-.289	.306
			MM_S3F	1.7 (1.0)	.8 (0.8)	15 ***	.889	.501	1.429
			MM_SRE	634(82.5)	770(62.0)	12 ***	1.917	1.361	2.939
	2005/06	40	MM_S1R	4.3 (1.5)	6.2 (1.6)	30.5 ***	1.215	.761	1.916
			MM_S1F	2.9 (1.5)	2.8 (1.3)	182	.056	-.406	.264
			MM_S3F	2.4 (1.1)	1.4 (1.0)	75 ***	.777	.399	1.308
			MM_SRE	955(172.7)	1160(123.1)	17 ***	1.320	.873	2.003
Test	**Cohort**	**N**	**Variable**	**Pre [V][a]**	**Post [N][a]**	**t_c[b]**	**d_{δ}[c]**	**$CI_{95\%}$ d_{δ}**	
								lower	upper
Problem Solving Knowledge Test	2003/04	66	V/N _5	12.7 (2.6)	13.9 (2.5)	3.16 **	.367	.119	.675
			V/N_5_{resp}[e]	18.8 (2.5)	19.1 (2.1)	.83	.006	-.170	.315
			(5-5_{pec})/5_{pec}[f]	.209(.106)	.258(.102)	3.25 **	.338	.108	.618

*p<.05, **p<.01, ***p<.001 (one-tailed).

[a]M (SD).

[b]Wilcoxon matched-pairs signed-ranks T test statistic.

[c]Effect size estimates (Cohen's *d*) and confidence intervals based on ordinal δ statistic (Cliff, 1993, 1996; Long et al., 2003). The signs of the effect size statistics were adjusted; positive values always indicate learning gains.

[d]Student t for dependent samples test statistic.

[e]Total number of attempted responses, i.e. item count without "don't know"-responses.

[f]Chance corrected sum score (based on maximum likelihood probability mean score; see text).

The Problem Solving Knowledge sum scores have also increased significantly between pre-test [V] and post-test [N] ($\Delta M_{N\_5\text{-}V\_5}$=1.205, t_c(65)=3.161, p=.001 one-tailed with Cohen's d_{t_c}=.474 with a 95% CI of .169 to .779; ordinal δ_b=.254 p=.001 with Cohen's d_{δ_b}=.367). Hypothetically, this difference in raw sum scores could be influenced by the number of items

actually responded to, however, as for each item the "don't know" option could be selected by participants. The analysis shows, however, that the number of items actually answered to (V_5_{resp}, N_5_{resp}) has not significantly changed between pre- and post-test; neither in a within-subject comparison (Wilcoxon T(66)=725, Z_T= -.812, p=.417; δ_w= .005, p=.654) nor in the analysis of change in the group as a whole (interval t_c(65)=.830, p=.410 two-tailed with Cohen's d_{t_c}=.130 with a 95% CI of -.178 to .438; ordinal δ_b=.005, p=.300).

Nevertheless, to promote comparability between participant scores, the number of items responded to was taken into account when testing if the sum of scores are actually higher than what could have been expected by randomly selecting item answers. In both pre-test [V] and post-test [N], the differences were calculated between the total sum scores (V_5 and N_5) and the expected sum scores (V_5_{pec}, N_5_{pec}), divided by the number of items actually responded to. The difference between pre- and post-test in this measure results in an index comparable to what Varga and Delaney (2000) have termed the stochastic superiority measure. Here, medium increases are found as well (t_c(65)=3.252, p=.001 one-tailed with Cohen's d_{t_c}=.474 with a 95% CI of .169 to .779; ordinal δ_b=.394, p=.0010; δ_w=.236, p=.0003).

DIGEST 4.1

Three types of indicators for learning are distinguished.

First, in module online surveys, participants were asked to report learning gains in pre-defined areas. The scores allowed for estimating the achievement of potential learning objectives per module.

Second, self-ratings as to the scope and confidence in one's knowledge and skills were solicited repeatedly across time. Results indicate that changes in what is referred to as 'self-efficacy' levels are essentially linked to particular PROCON modules- In terms of Cohen's (1988) classification, the self-efficacy gains can be considered "medium" to "large" effects.

And, third, knowledge and situational judgement test scores are available to evaluate the extent of learning. Significant differences between pre- and post-test knowledge scores were found (Cohen's d≈+.3 to .4). As for situational judgement, large differences between pre- and post-tests were found for scores based on 'correct' answers (d≈+1.2 to +2.5) and those denoting the occurrences of salient errors (d≈–.8).

4.2 Indicators of Module Preparation

Indicators of CSSL Courseware Use.

The *Conflict Resolution in Groups* courseware was introduced in 2003, and in both 2004 and 2005, all modules were delivered in a hybrid format.

Table 21

Indicators of Courseware Use in Computer-Supported Social Learning Modules

Cohort / Module / Variable							**[Time]→**		
2003/04							**Confl. Res. in Groups**		
Participants [n]							n=79[a]		
Degree of use [1-high, 6-low]							[K04b_01]	1.9	(1.3)
Degree of use [% used]							[K02_022]	86.2	(24.0)
Time spent [h]							[K02_014]	10.5	(6.1)
Log-file reports [n]							n=53		
Page completion							[L_ACPAGE]	47.9	(22.7)
Task completion							[L_TASKC]	68.5	(31.3)
Task input quantity							[L_TASKL]	8.4	(5.0)
Sum of input lengths							[L_CHARS]	4715	(2565)
Overall rating [15(best)…0pt]							[K03_13]	10.4	(2.1)
2004/05	**Problem Solving**			**Mediation Skills**			**Confl. Res. In Groups**		
Participants [n]	N=42			n=43			n=43		
Degree of use [1-high, 6-low]	[P02_01]	2.6	(1.7)	[M02_01]	2.3	(1.4)	[K04b_01]	2.5	(1.6)
Degree of use [% used]	[P02_064]	79.2	(28.4)	[M02_082]	85.5	(25.0)	[K02_023]	80.7	(27.5)
Time spent [h]	-			[M02_094]	8.7	(3.9)	[K02_014]	13.1	(6.1)
Time required [h]	[P02_074]	19.9	(7.7)	[M02_095]	13.4	(4.7)	[K02_015]	18.4	(5.3)
Log-file reports [n]							n=36		
Page completion							[L_ACPAGE]	67.6	(25.8)
Task completion							[L_TASKC]	63.1	(24.1)
Task input quantity							[L_TASKL]	25.8	(15.5)
Sum of input lengths							[L_CHARS]	17129	(10533)
Overall rating [15(best)…0pt]	[P04_01]	10.3	(2.4)	[M04_01]	10.6	(2.3)	[K03_13]	10.1	(2.3)
2005/06							**Confl. Res. In Groups**		
Participants [n]							n=43		
Degree of use [1-high, 6-low]							[K04b_01]	3.2	(1.5)
Degree of use [% used]							[K02_023]	65.8	(33.9)
Time spent [h]							[K02_014]	7.2	(5.4)
Time required [h]							[K02_015]	13.0	(7.1)
Overall rating [15(best)…0pt]							[K03_13]	10.7	(1.3)

Note Data are arithmetic mean (standard deviation) unless noted otherwise. All Information is based on Log-file Analyses (see Table 15) and Module Online Surveys for Problem Solving [P], Mediation Skills [M], and Conflict Resolution in Groups [K] as detailed in Appendix B. [a]1 participant did not fill out the module online survey.

The indicators of courseware use as listed in Table 21 show that, in general, a majority of participants reports to have worked with the coursewares. The most common value reported as to what extent (i.e. percentage) the contents of the respective courseware had been worked through is Mo=100% (see also Table 21 above), and this applies to all CSSL modules and

online surveys [P02_064, M02_083, K02_022/023]. Likewise, all participants of the blended learning cohorts were asked to grade their exposure to the coursewares on a 1 through 6 Likert scale (with "1" meaning that they had thoroughly followed up all aspects of the courseware and "6" signifying that that they had not used the courseware at all [P02_01, M02_01, K04b_01]). In the 2003/04 and 2004/05 cohorts the mode of these variables was also Mo=1, and in the 2005/06 cohort it was 2.

Validity of Courseware Use Measures.

The quality of these self-reported data shows when it is compared to the empirical findings from the log-file analyses. To determine whether the self-reported indicators of courseware can be used as valid criteria, they were matched with the empirical indicators based on the log-file analyses (see Table 16). The results are reported in Table 22.

Table 22

Correspondence between Self-reported and Empirical Indicators of Courseware Use

***Conflict Resolution in Groups* Module**	**2003/04[a]**		**2004/05[b]**	
r_s Degree of use [K04b_01, 1-6] x Degree of use [K02_022/023, %]	-.714	***	-.772	***
r_s Degree of use [K04b_01, 1-6] x Page completion [L_ACPAGE]	-.472	***	-.526	***
r_s Degree of use [K04b_01, 1-6] x Task completion [L_TASKC]	-.376	**	-.526	***
r_s Degree of use [K04b_01, 1-6] x Task text input quantities [L_TASKL]	-.349	**	-.675	***
r_s Degree of use [K04b_01, 1-6] x Sum of input lengths [L_CHARS]	-.388	**	-.647	***
r Degree of use [K02_022/023, %] x Page completion [L_ACPAGE]	.824	***	.802	***
r Degree of use [K02_022/023, %] x Task completion [L_TASKC]	.796	***	.802	***
r Degree of use [K02_022/023, %] x Task text input quantities [L_TASKL]	.598	***	.694	***
r Degree of use [K02_022/023, %] x Sum of input lengths [L_CHARS]	.678	***	.684	***
R Degree of use [K02_022/023, %]. [L_ACPAGE] [L_TASKL]	.840 [c]	***	.822 [d]	***
r Hours used [K02_014, h] x Page completion [L_ACPAGE]	.224	***	.609	***
r Hours used [K02_014, h] x Task completion [L_TASKC]	.218	***	.609	***
r Hours used [K02_014, h] x Task text input quantities [L_TASKL]	.276	***	.708	***
r Hours used [K02_014, h] x Sum of input lengths [L_CHARS]	.272	***	.698	***
R Hours used [K02_014, h]. [L_ACPAGE] [L_TASKL]	.288 [c]	***	.796 [d]	***

* p < .05; ** p < .01; *** p < .001 (two-tailed).

Note. $r_{L\_ACPAGE\&L\_TASKL}$(2003/04)=.561, r(2004/05)=.714.

[a]n=54. [b]n=36. [c]R^2=.706, F(2,50)=60.15, p<.000. [d]R^2=.676, F(2,32)=33.41, p<.000.

The absolute values of the correlation coefficients listed in Table 22 are generally higher for the self-reported percentage indicators than for the self-reported degree of use on the 1…6 Likert scale. They are also generally higher in the 2004/05 cohort, especially for the self-reported number of hours in which participants have prepared themselves with the coursewares [K02_014]. This is probably due to improved and more detailed log-filing procedures.

The values of the correlation coefficients between the self-reported degree of use (on a 0-100% scale) and the log-files range between r≈.60 and .82 and are highest in the page completion category [L_ACPAGE], a measure of *breadth* or *coverage* of courseware use. As for *depth* of courseware use, mainly represented by the average quantities of text entered across courseware input tasks [L_TASKL], the correlation with self-reported percentage indicator is somewhat lower with r≈.60 and .69, respectively.

Also, the multiple correlation between the self-reported percentage and the empirical indicators is only marginally higher (R=.84 and R=.82, respectively) than the single r between the percentage and the page completion index (r=.82 and .80), probably due to high correlation between the two predictors.

The results are generally in line with Kilburg's (2005b, p. 69) finding who calculated a (possibly somewhat over-estimating) Pearson correlation of r=.86 (p = <.001; n= 32) between the self-reported degree of use [P02_064] and a log-file indicator for the 2004/05 *Problem Solving* module.

In summary, it does not seem unreasonable to assume that the self-reported variables are generally is a valid indicator of courseware use across modules and cohorts.

Courseware Use Indices for Impact Analysis.

It was therefore decided to use the self-reported variables with an interval level-of-measurement as indicators of the actual utilization of computer-supported learning elements by students.

As Table 23 indicates, however, most participants actually did work with the courseware offered (which is no surprise given the fact that the production of the courseware originally resulted from a need voiced by learners). For multivariate impact analysis, however, measures are needed which are, on the one hand, approximately normally distributed. On the other hand, using ex-post facto comparison groups as a basis for analysis, major problems could possibly arise from small expected frequencies; thus, a certain minimum of participants in each subgroup is needed.

Table 23

Self-Reported Conflict Resolution in Groups *Courseware Use In Cohorts*[a]

Number of Survey Respondents (n) in Cohort [Variable]	0%	10	20	30	40	50	60	70	80	90	100%	Σ	M	(SD)
2003 [K02_022]		2	2	2	1	2	4	3	7	6	50	**79**	**86.2**	**(23.98)**
2004 [K02_023]	1		3	1		3	1	3	2	10	19	**43**	**80.7**	**(27.46)**
2005 [K02_023]	2	3	1	4	2	2	4	2	2	5	16	**43**	**68.4**	**(34.15)**
Total	**3**	**5**	**6**	**7**	**3**	**7**	**9**	**8**	**11**	**21**	**85**	**165**	**80.1**	**(28.63)**

[a]Extent to which the contents of the respective courseware had been worked through (response set options: 0-100%, in 10% steps). Cells contain number of survey respondents (n).

In order to produce an index which is it seemed a worthwhile effort to combine the information contained in the three continuous variables (percentage degree of use to start of module [K02_022], percentage of use to date [K02_023] and hours used [K02_014]), while simultaneously correcting for characteristics only present in specific cohorts. This index was calculated by using the z-transformed indices per cohort and retransforming them so that (a) no negative values exist and (b) the re-transformed distribution has a minimum of either 0 or a maximum of 100, depending on the highest absolute z value. As the validities of the individual indices were essentially unknown, it was decided to use both to an equal weight in a geometric mean. Thus, the courseware usage index (CUI) was calculated per blended-learning cohort as

$$CUI_K = \sqrt{\left(\frac{50z_{[K02\_023]}}{\max(|\min(z_{[K02\_023]})|,|\max(z_{[K02\_023]})|)}+50\right)\left(\frac{50z_{[K02\_014]}}{\max(|\min(z_{[K02\_014]})|,|\max(z_{[K02\_014]})|)}+50\right)}.$$

To allow for analyses of the combined courseware deployment, an index was accordingly calculated for the 2004/05 *Mediation Skills* blended learning module as

$$CUI_M = \sqrt{\left(\frac{50z_{[M02\_082]}}{\max(|\min(z_{[M02\_082]})|,|\max(z_{[M02\_082]})|)}+50\right)\left(\frac{50z_{[M02\_094]}}{\max(|\min(z_{[M02\_094]})|,|\max(z_{[M02\_094]})|)}+50\right)}.$$

For resulting CUI values <1 it was assumed that courseware had not been used at all, and, accordingly, these cases were treated equivalent to the non-blended learning cohort.

Moreover, to express usage of all conflict- and mediation related coursewares across all comparative cohorts, an integrative index CUI_{TTL} was used with $CUI_{TTL}=0$ for the 2002/03 (non-blended-learning) cohort, with $CUI_{TTL}=CUI_K$ for the 2003/04 cohort. For the 2004/05 cohort, usage of the *Mediation Skills* courseware was added as an additional component with

$$CUI_{TTL}^{2004/05} = \sqrt{\left(50\frac{[K02\_023]+[M02\_094]-M_{[K02\_023]}}{s_{[K02\_023]}}{\max(|\min(z_{[K02\_023]})|,|\max(z_{[K02\_023]})|)}+50\right)\left(50\frac{([K02\_014]+.5[M02\_082])'-M_{[K02\_014]}}{s_{[K02\_014]}}{\max(|\min(z_{[K02\_014]})|,|\max(z_{[K02\_014]})|)}+50\right)},$$

where the z calculations include the self-reported number of hours spent on the *Mediation Skills* courseware [M02_094] and the self-reported percentage degree of use of the *Mediation Skills* courseware [M02_082]. Where the term ([K02_014]+.5[M02_082]) was greater than 100(%), 100% was assumed so as not to unnecessarily distort the index.

Table 24

Courseware Usage Index (CUI_K) for Blended Conflict Resolution in Groups Modules

	2003/04[a]		2004/05[b]	
r_s CUI_K x Degree of use [K04b_01, completely:1-not at all:6]	-.531	***	-.707	***
r CUI_K x Percentage used [K02_022/23, 0-100%]	.808	***	.885	***
r CUI_K x Hours used [K02_014, h]	.742	***	.898	***
r CUI_K x Page completion [L_ACPAGE]	.655	***	.767	***
r CUI_K x Task completion [L_TASKC]	.637	***	.767	***
r CUI_K x Task text input quantities [L_TASKL]	.527	***	.817	***
r CUI_K x Sum of input lengths [L_CHARS]	.582	***	.810	***

* p < .05; ** p < .01; *** p < .001 (two-tailed). [a]n=77 for within self-report [K], n=54 for empirical indicators. [b]n=42 within self-report [K], n=36 for empirical indicators.

When comparing the correlations between the indicators of courseware use with the empirically found indices listed in Tables 23 and 24, the CUI estimate compares favourably. This could indicate that this model the combination of information contained in the CUI actually is a better predictor of empirical indicators of courseware use; it may also mean, however, that this model may be more apt for parametric statistical analyses with its underlying unimodal distributions, comparatively minor problems of somewhat negative skewness and kurtosis, and without salient violations of the normality assumption (see Fig. 22a through c for examples).

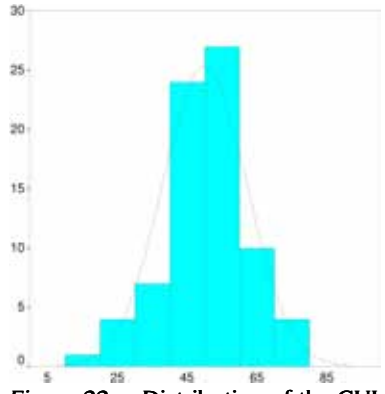


Figure 22a. Distribution of the CUI (2003/04).

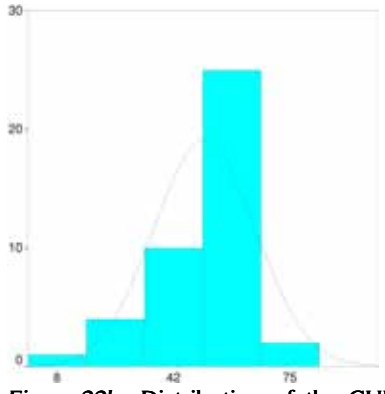


Figure 22b. Distribution of the CUI (2004/05).

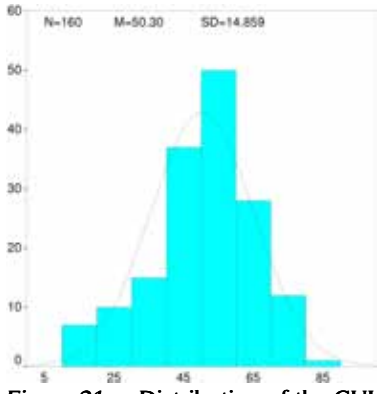


Figure 21c. Distribution of the CUI (all blended learning cohorts).

In no single case, the null hypotheses of the Kolmogorov-Smirnov goodness-of-fit test that the samples are derived from normally distributed populations could be rejected. Table 25 below lists characteristics of all Courseware Usage Indices (CUI) calculated.

Table 25

Characteristics of the Courseware Usage Indices (CUI) calculated

	Conflict Resolution in Groups [CUI_K]					*Med.Skills* [CUI_M]	Composite [CUI_{TTL}]
	2003/04	2004/05	2003-05	2005/06	2003-06	2004/05	2004/05
N	80	43	123	43	166	40	41
- N thereof with CUI<1	3	1	4	2	6	2	0
M	49.97	50.51	50.16	50.70	50.30	51.13	97.60
Median (based on grouped data)	50.63	51.94	50.96	54.55	51.34	53.37	105.53
SD	12.078	14.594	12.963	19.539	14.859	12.411	24.268
K-S max|d|[a]	.137	.139	.113	.099	.086	.107	.161
K-S Z[b]	1.199	.900	1.231	.635	1.082	.675	1.034
K-S p[c]	.113	.393	.096	.815	.193	.752	.235
Ex-Post Facto Grouping	K_w	K_w	K_{B1}	K_w	K_{B2}	M_w	CUI_B
[0] LOW USAGE: CUI between	0-50.6	0-53.0	0	0-50.0	0-23.0	0-53.0	0-23.5
[1] MEDIUM	-	-	>0-50.68	-	>23-52.16	-	>23.5-56.9
[2] HIGH USAGE	>50.6	>53.6	>50.68	>50.0	>52.16	>53.0	>56.9

* p < .05; ** p < .01; *** p < .001 (two-tailed). [a]Highest absolute difference between distributions in Kolmogorov-Smirnov Normality Goodness-of-Fit Test. [b]Kolmogorov-Smirnov Normality Goodness-of-Fit Test Z-value. [c]Kolmogorov-Smirnov Normality Goodness-of-Fit Test p-value.

For purposes of within-cohort comparisons (K_w, M_w) as well as for between-group comparisons (K_{B1}, K_{B2}, CUI_B), CUI indices were also used to subdivide the entire sample base into ex-post facto comparison groups by means of median and tercile split criteria. As indicated in Table 25, these criteria were adjusted, where necessary, to arrive at approximately equal subsample sizes.

In order to find possible predictors of courseware use and to check whether differences exist at pre-test, cohort-based Mann-Whitney U tests were performed for the pairs of K_w and M_w groups for all relevant pre-test variables (i.e. socio-demographic background, self-efficacy, knowledge and situational judgement test sum scores). No significant differences were found between groups and across cohorts using a stepwise Bonferroni-correction as suggested by Larzelere & Mulaik (1977) to determine critical values for the $|Z_U|$ based exact p-levels. As this is a relatively conservative approach possibly lacking power, a second methodology to determining possible predictors of courseware use was used. Here, Mann-Whitney $|Z_U|$-scores and exact p-levels for each variable were averaged across the k blended learning cohorts, and allowing the family-wise error rate to be $k\alpha$. Applying this empirical approach to the *Conflict Resolution in Groups* courseware groups K_w yields three variables with consistent significant differences across

the three blended learning cohorts, namely, [V_SE03], the self-rated facilitation knowledge ($M_{|Z_U|}$=1.405, M_p=.096), and two salient error sum scores in the *Facilitation Knowledge* subtest [V_Mod1F] ($M_{|Z_U|}$=1.311, M_p=.146) and [V_Mod3F] ($M_{|Z_U|}$=2.033, M_p=.031). In all cases, lower levels of knowledge seem to be associated with higher courseware use (Fisher-Z-averaged mean r_s across all weighed groups and the three variables is -.208). In addition, an analysis of pre-test equivalence of ex-post facto groups in terms of demographical variables was performed to identify possible confounding factors. As detailed in Appendix D.2, systematic between-group differences were found for two variables, namely the Duration of Study (with a tendency for senior students to report lower courseware use), and, interestingly, the Sensitivity (*Sensibilität*) score on the standardized *BIP* subtest (with a trend for more "sensitive" participants in terms of the *BIP* to report higher courseware use).

In addition to the ex-post facto group-split based on the CUI indices, however, and for purposes of improved internal validity, it seemed sensible to follow an additional approach to creating ex-post facto groups directly based on the self-reported degree of use. As for this approach based on Table 23 above, it can reasonably be argued above that the sole use of empirical criteria to split the groups – which could help steering clear of statistical problems possibly arising from unequal sample sizes – is barred on logical grounds. For example, theoretical grouped data percentiles for the total sample amount to P_{25}=66.8, $P_{33.3}$=80.9, P_{50}=80.9 and $P_{66.6}$=97.6. It seemed sensible, therefore, not to mesh participants who did hardly or not use the courseware with those possibly profiting from using it to an extent of 60 or even 80 percent.

To balance these contradictory objectives, it was decided, on the one hand, to include at least n=5 participants in each cohort subgroup. On the other hand, it seemed reasonable that the first category should not include values higher than 40% to meet the objective that the subgroups represent different degrees of courseware utilization. Given these restrictions, the 10 resulting alternatives for grouping were assessed in terms of the objectives enunciated above. Table 26 details the structures of the various alternatives. All alternatives were tested for homogeneity across subgroups x cohorts by using the χ^2-likelihood-ratio as an index. Association between the ordinal subgroups and the reported percentages (Kendall's τ-c), and validity indices based on subgroup affiliation predicting courseware use (Goodman and Kruskal's Lambda and Tau) are also listed in Table 25.

While a four-group split generally achieves higher values for association and predictive accuracy than three-group splits, the heterogeneity index also tends to be higher.

Furthermore, detailed comparisons of the alternatives listed in Table 26 indicate that the alternative with the best trade-off between homogeneity and predictive accuracy is B. The latter also displays the highest absolute ordinal correlation between subgroup ranks and the self-reported degree of use on the 6-point Likert scale percentages (see Table 27). Here, other subgroup-split-ups seem to be associated more strongly with the empirical indices; however, these generally tend to display lower absolute correlations with the self-report indices. It was therefore decided to use Alternative B for subsequent ex-post facto between-group analyses.

Table 26

Ex Post Facto Group Alternatives Based On Courseware Use Level

	Subgroups				Goodness-of-Fit							
					Blended Learning Cohorts[a]				All Cohorts[b]			
Alt	1	2	3	4	LR[c]	τ_C[d]	λ[e]	τ_{GK}[f]	LR	τ_C	λ	τ_{GK}
A	0-30%	40-70%	80-100%		9.6	.68	.200	.264	191.8	.88	.537	.426
B	0-30%	40-70%	80-90%	100%	13.1	.87	.463	.575	167.7	.91	.680	.631
C	0-30%	40-80%	90-100%		7.9	.77	.225	.323	190.0	.93	.551	.465
D	0-30%	40-80%	90%	100%	14.8	.87	.488	.613	197.0	.92	.694	.656
E	0-30%	40-90%	100%		11.3	.87	.350	.470	193.5	.99	.619	.562
F	0-40%	50-70%	80-100%		10.4	.68	.200	.266	184.3	.87	.537	.419
G	0-40%	50-70%	80-90%	100%	13.9	.87	.463	.577	187.8	.91	.680	.624
H	0-40%	50-80%	90-100%		8.9	.78	.225	.324	182.8	.92	.551	.457
I	0-40%	50-80%	90%	100%	15.8	.87	.488	.614	189.8	.91	.694	.649
J	0-40%	50-90%	100%		13.0	.90	.350	.472	186.8	.98	.619	.555

[a]Cohorts 2003/04, 2004/05, and 2005/06, n=165. [b]All cohorts, with 2002/03 in subgroup 1 and 0% as no courseware utilization; n=232. [c]Likelihood Ratio (LR) for cohort x subgroups contingency tables with *df*=4 for three and *df=6* for 4 subgroups and the three blended learning cohorts (2003/04, 2004/05, and 2005/06), and for all cohorts with *df=6* or *df=9*, respectively. [d]Kendall's Tau-C (τ) for symmetric association between subgroups x (original) self-reported courseware percentage degree of use [K02_022/023, %]. [e]Goodman and Kruskal's lambda for self-reported courseware degree of use [K02_022/023, %] predicted. [f]Goodman and Kruskal's tau for self-reported courseware degree of use [K02_022/023, %] predicted.

Table 27

Ex Post Facto Group Alternatives: Tau-b-Correlation with Courseware Use Indices

Alt	Subgroups				Courseware Use Indices			
	1	2	3	4	Percentage used (self-reported)[a]	Degree of use (self-reported)[b]	Depth (empirical) [L_TASKL][c]	Breadth (empirical) [L_ACPAGE][d]
A	0-30%	40-70%	80-100%		.881	-.619	.378	.452
B	0-30%	40-70%	80-90%	100%	.953	-.675	.283	.368
C	0-30%	40-80%	90-100%		.905	-.634	.349	.456
D	0-30%	40-80%	90%	100%	.952	-.663	.280	.373
E	0-30%	40-90%	100%		.931	-.657	.276	.350
F	0-40%	50-70%	80-100%		.875	-.615	.378	.452
G	0-40%	50-70%	80-90%	100%	.948	-.671	.283	.368
H	0-40%	50-80%	90-100%		.900	-.632	.349	.456
I	0-40%	50-80%	90%	100%	.948	-.661	.280	.373
J	0-40%	50-90%	100%		.928	-.658	.276	.350

[a]Extent to which the contents of the respective courseware had been worked through (response set options: 0-100%, in 10% steps). Based on n=232 respondents (all cohorts). [a]Response set on 6-point Likert scale with 1-"completely" through 6-"not used". Based on n=165 (blended-learning cohorts). [c]Based on n=89 analyses. [d]Based on n=88 analyses.

Use of Preparational Materials.

As detailed above, the peer-tutored *PROCON* classroom training sessions had a strong focus on the development of practical skills and shared process knowledge. By contrast, the delivery of conceptual knowledge is traditionally focused on in books or other written materials (which, naturally, may include images, hypertexts, etc.). The CSSL coursewares were introduced into the *PROCON* curriculum with the idea to enhance (not to replace) the traditional delivery of conceptual knowledge to better prepare students for the training courses. By contrast, there was some anecdotal evidence in the feed-back from peer tutors that, after the introduction of coursewares, participants tended to read and study written materials to a lesser degree. However, no significant changes were found upon analysis of the self-reported degrees of text-book use as depicted in Figure 23 below. The only significant change across time found is a lower degree of concomitant further reading for the *Conflict Resolution in Groups* module in the 2005/06 cohort. Compared to the previous cohort 2004/05, 57.2% of the scores are higher, indicating that reading recommendations were less received, as compared to 20.7% for the reverse, resulting in a significant Cliff's δ=(57.2%-20.7%=).365 (Cohen's d_δ=.566 with a 95% CI ranging from .180 to 1.011).

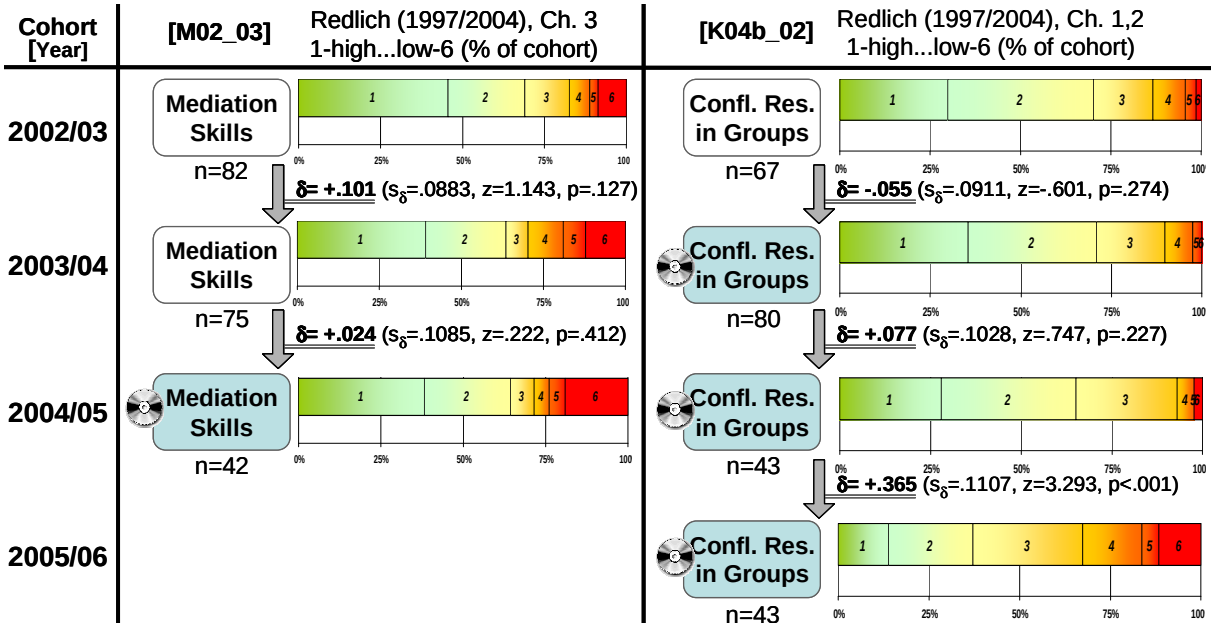


Figure 23. **Variations in self-reported degrees of text book penetration in mediation-related modules.**

In order to obtain a more comprehensive picture, "textbook" or "literature" usage indices were calculated, comprising not just self-reported use of the main textbook as indicated in Figure 23, but the self-reported use of all preparatory reading consistently recommended across the various cohorts. Definitions for the exact composition of the reading materials usage indices for the *Mediation Skills* [M_Lit] and *Conflict Resolution in Groups* modules [K_Lit] are detailed in Appendix B.

Figures 24 and 25 depict the degrees of module preparation in terms of textbook penetration in the respective cohorts. In terms of both intensity and extent of textbook use per module, there appear to be sustained declines after the introduction of the coursewares, and this seems to apply to both the *Conflict Resolution in Groups* and the *Mediation Skills* modules. By contrast, however, positive within-cohort associations were found between textbook and courseware use for the 'blended' *Conflict Resolution in Groups* modules. These have also been found to decline across subsequent cohorts, nevertheless, and a negative, albeit not significant, association has been found in the 2004/05 blended learning *Mediation Skills* module.

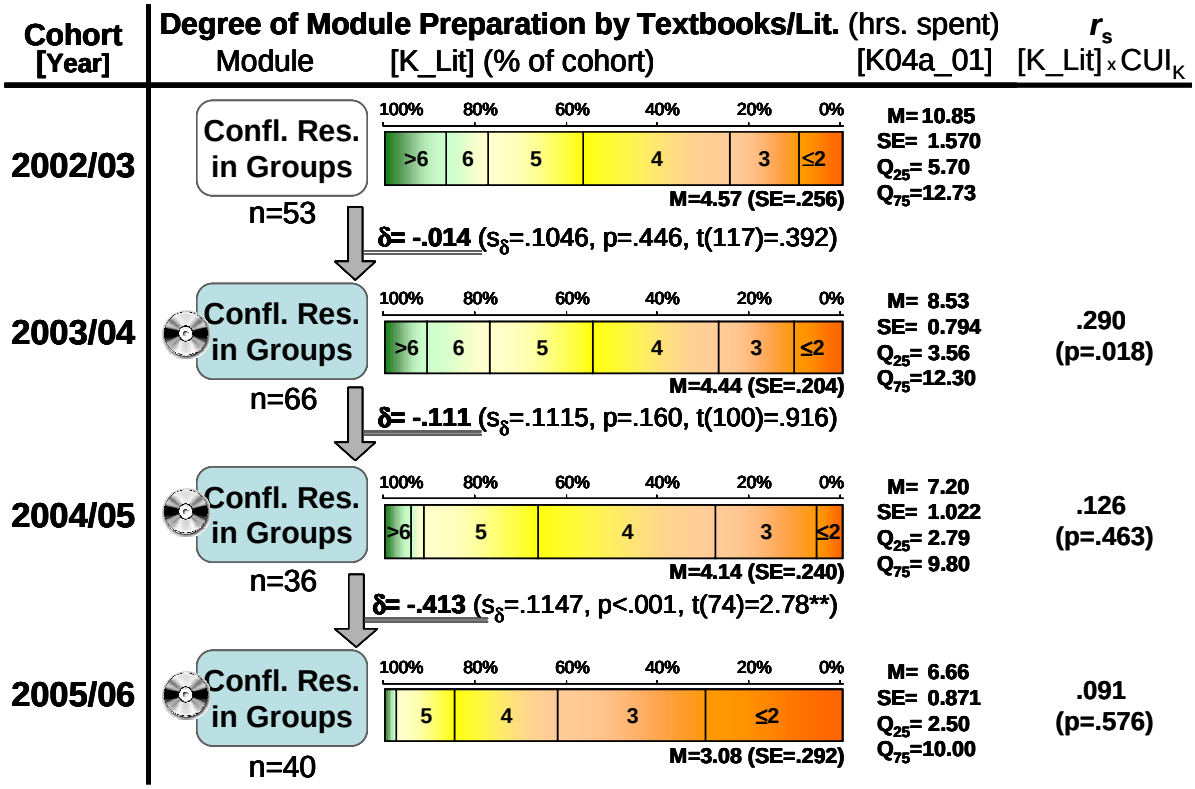


Figure 24. Intensity and duration of preparation for the PROCON Conflict Resolution in Groups module by use of textbooks/Literature (by Cohort).

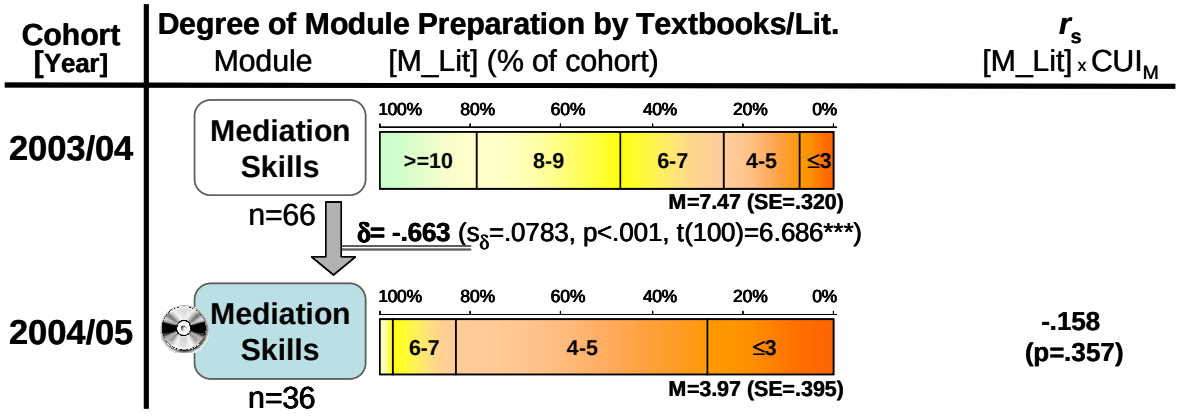


Figure 25. Intensity and duration of preparation for the *PROCON Mediation Skills* module by use of textbooks/literature (by cohort).

Analogous to the courseware use indices, the textbook preparation indices were also used to divide the sample base into ex-post facto comparison groups based on criteria detailed in Table 28 below. Again, criteria were adjusted where necessary to arrive at approximately equal subsample sizes.

Table 28

Characteristics of the Textbook/Literature Preparation Indices

	2002/03	2003/04	2004/05	2002/03–2004/05	2005-06	All Cohorts	
N	53	66	36	155	40	195	
***Conflict Resolution in Groups* [K_Lit]**							
M	4.57	4.44	4.14	4.41	3.08	4.14	
Median[f]	4.36	4.36	4.09	4.29	3.14	4.05	
SE	.256	.204	.240	.135	.292	.129	
K-S max|d|[a]	.185	.150	.205	.178	.184	.167	
K-S Z[b]	1.349	1.219	1.231	2.211	1.162	2.327	
K-S p[c]	.053	.102	.097	<.001	.134	<.001	
***Mediation Skills and Conflict Resolution in Groups* (combined) [KM_Lit][d]**							
M	10.06	11.91	8.11	10.39	–	–	
Median[f]	9.50	11.69	8.45	9.95	–	–	
SE	.605	.466	.476	.328	–	–	
K-S max|d|[a]	.170	.095	.126	.119	–	–	
K-S Z[b]	1.237	.770	.759	1.481	–	–	
K-S p[c]	.094	.593	.613	.025	–	–	
Ex-Post Facto Groups	LIT_W	LIT_W	LIT_W	LIT_{B1}	LIT_W[e]	LIT_{B2}[e]	LIT_{B3}[g]
[0] LOW USE: [KM_Lit]	0-9	0-11	0-8	0-8	0-3	0-3	0-9
[1] MEDIUM	–	–	–	9-11	–	4	–
[2] HIGH USE	>=10	>=12	>=9	>=12	>=4	>=5	>=10

*p < .05; **p < .01; ***p < .001 (two-tailed). [a]Highest absolute difference between distributions in Kolmogorov-Smirnov Normality Goodness-of-Fit Test. [b]Kolmogorov-Smirnov Normality Goodness-of-Fit Test Z-value. [c]Kolmogorov-Smirnov Normality Goodness-of-Fit Test p-value. [d][KM_Lit]=[K_Lit]+[M_Lit] (see Appendix B.) [e]Based on the [K_Lit] index. [f]Based on grouped data. [g]For the 2005/06 cohort, LIT_{B3} was based on the [K_Lit] index with [K_Lit]<4 constituting the "LOW" group and participants with values equal to or above 4 classified into the "HIGH" group.

DIGEST 4.2

The mediation-related CSSL coursewares which this study seeks to investigate were introduced into the *PROCON* curriculum to enhance the 'traditional' module preparation by use of textbooks and other written and printed materials. Therefore, indices were defined for both courseware use and the employment of recommended reading materials by participants.

Descriptive analyses reveal that, across the blended leaning cohorts, a majority of participants reports to have worked with the coursewares to a significant degree. As these self-reports were found in general accordance with the data empirically stored in log-files during courseware use ($r \approx +.6$ to $+.8$) , they were used as a basis for calculating courseware use indices (CUI) to be used for impact analysis. Moreover, for between-group comparisons (e.g. 'low' vs. 'high' courseware use), participant subgroups were identified both for within-cohort and across-cohort analyses.

In the same fashion subgroups were defined based on the indices denoting both intensity and extent of textbook use for preparation of mediation-related training courses. There appear to be significant declines in textbook use after the introduction of the *Mediation Skills* courseware (2004/05) and in comparison between the 2005/06 and previous cohorts (Cohen's $d \approx -.4$ to $-.6$).

4.3 CSSL and Learning

In the following, the evaluative questions posed above will be examined using straightforward statistical procedures. Where appropriate, both measures of correspondence and difference will be reported based on both parametric and distribution-free tests to advance the internal validity of conclusions drawn.

H1: CSSL and Situational Judgement.

The first question to be examined is whether courseware use may correspond with superior situational judgment. It was detailed above how courseware usage is taken to be measured by the respective indices. It was also detailed above that all blended learning cohorts were given versions of the situational judgement test (SJT) at post-test [N] with parallel versions used between 2003 and 2005. Despite the fact that all post-test SJT sum scores were found to be unimodal and symmetric, the normality assumption was rejected in the Kolmogorov-Smirnov goodness-of-fit tests, saliently at times. Therefore, where possible, non-parametric alternatives to the traditional statistical procedures were used in the following.

In a first approach to the question posed, Spearman rho coefficients were calculated; the following Table 29 gives an overview.

Table 29

Association between Courseware Usage and Post-Test Situational Judgement Sum Scores

					CUI_{TTL} Effect Size Estimate		
						95% CI *d*	
SJT Sum Score		CUI_K	CUI_M	CUI_{TTL}	Cohen's *d*	lower	upper
N_MM_S1R	r_s	.395 ***	.300 ***	.449 ***	1.005	.648	1.389
N_MM_S1F	r_s	-.029	.052	-.017	-.035	-.366	.296
N_MM_S2R	r_s	.245 **	.128	.234 **	.481	.147	.828
N_MM_S2F	r_s	-.067	.058	.007	.013	-.318	.345
N_MM_S3F	r_s	-.081	-.253 **	-.179	-.364	-.705	-.032
N_MM_SRE	r_s	.187 *	.097	.189 *	.385	.053	.727

* p < .05; ** p < .01; *** p < .001 (one-tailed). N=153. Sample base cohorts included in this analysis were 2002/03 (no blended learning), 2003/04, and 2004/05. 2005/06 was not included as a revised version of the post-test had been used. Effect size and confidence interval estimates based on Cohen's $d=2r(1-r^2)^{-1/2}$. The CI values are based on Fisher-Z-transformation for Spearman's r as detailed in Zar (1999; cit. in Sheskin, 2007, p. 1364) with $CIz_{(95\%)} = 0.5 \ln[(1+r)/(1-r)] \pm 1.96 [1.060/(n-3)]^{1/2}$.

In the analysis of differences between pre- and post-test SJT scores (cf. Table 20), no significant changes were observed for salient error sum scores MM_S1F and MM_S2F. For the same variables, no significant association was found here. As for the other variables, however, courseware use seems to be positively associated with higher SJT sum scores. In line with the hypotheses for evaluation, the associations between the SJT sum scores and usage indices for

the *Conflict Resolution in Groups* courseware are generally higher than those between SJT sum scores and the *Mediation Skills* module courseware.

Generally, the associations between the Courseware Use Indices and SJT post-test scores are considerably weaker when the non-blended learning cohort (2002/03) is excluded from the analysis. While, across blended learning cohorts, all rank correlation coefficients display signs indicating direction as hypothesized, coefficients have not been found to significantly differ from zero. Moreover, the findings are even less consistent when associations are compared between the two blended-learning cohorts 2003/04 and 2004/05 individually. This may indicate that difference-based analyses may be more appropriate and/or powerful than correspondence-based analyses for the evaluative question under review.

In 2006, a revised version of the SJT post-test had been used (with a rating of options instead of simple multiple-choice). Moreover, participants in the 2005/06 cohort (N=40) seem to have used the *Conflict Resolution in Groups* courseware differently, i.e. variability is much higher in this group (cf. Table 23 in section 4.2). The pattern of association found resembles the one detailed in Table 29; all rank correlation coefficients display signs indicating direction as hypothesized with $r_{s\ CUI_K x\ N\_MM\_S1R}$ =.422 (p=.003 one-tailed, Cohen's d with 95% CI .238<.932<1.729), $r_{s\ CUI_K x\ N\_MM\_S1F}$ =-.186 (p=.125), $r_{s\ CUI_K x\ N\_MM\_S3F}$ =-.273 (p=.044 one-tailed, Cohen's d with 95% CI -1.301<-.567<-.104), and $r_{s\ CUI_K x\ N\_MM\_SRE}$ =.397 (p=.006 one-tailed, Cohen's d with 95% CI .176<.864<1.648).

It seemed therefore sensible to base the further analyses of associations predominantly on this cohort. One question arising from the data presented above concerns the influence of pre-test [V] results on both courseware use and the associations found between courseware use and SJT post-test scores. Theoretically, pre-test scores may be a good predictor of both courseware use and post-test results. This may be the case, for example, if all three are dependent on motivation to participate. For purposes of examination, partial coefficients were calculated based on Spearman's rho coefficients (r_s). The results of this examination presented in Figure 26 indicate that

(1) no significant associations were found between pre-test scores and courseware use;

(2) significant associations were found between courseware use and SJT post-test scores in three of the four examined variables and with all coefficient signs indicating association as hypothesized;

(3) the findings are somewhat ambiguous results as to the relationship between pre- and post-test SJT scores; and

(4) the portion of total variability in SJT post-test scores predicted by courseware use is equal or even higher (!) when pre-test scores are controlled for.

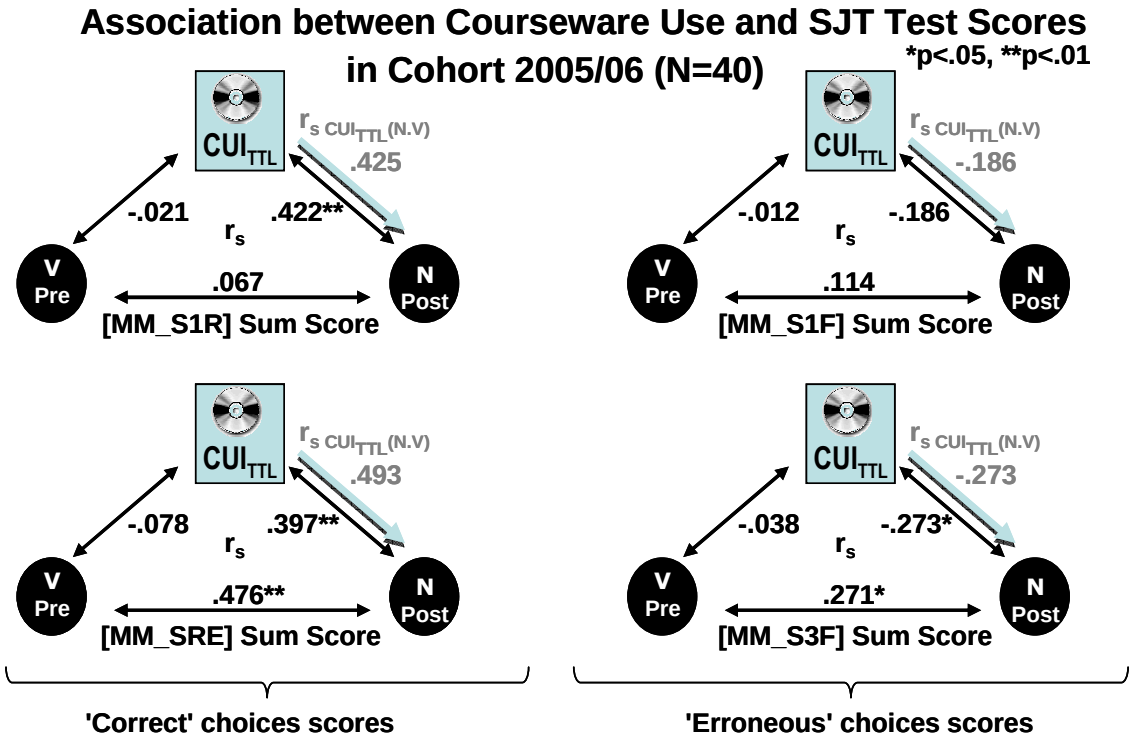


Figure 26.

In sum, these findings lend further support to the notion that that an association between courseware use and SJT post-test scores may exist. The association found is not very strong, however. At best, nearly a quarter of the variance found in SJT post-test sores may be uniquely predicted by courseware use ($r^2_{s\ CUI_{TTTL}(N.V)}$ for [MM_SRE]). Expressed in terms of a (traditional least-squares) linear regression estimate, approximately 5 score points of the post-test SJT sum score [MM_S1R] are achieved independent from usage of the *Conflict Resolution in Groups* courseware; an increase in the post-test SJT sum score by 1 point can then be achieved by nearly 8 hours of work or, alternatively, by using 56% of the courseware. (These figures should be treated with extreme caution due to the likely violations of assumptions of the underlying statistical model, however.)

To test which of these variables seem more important, a dominance-based ordinal multiple regression (DOMR) was performed, using methods developed by Cliff (1996a, 1996b, 1996c) and Long (1998, 1999, 2005; Long et al., 2003). The post-test SJT sum score with the

highest associations with self-reported measures of courseware use (N_MM_S1R) was used to estimate DOMR weights for the variables "percentage of use" of the two coursewares (*Mediation Skills* and *Conflict Resolution in Groups*) as well as for the reported number of hours spent learning with these coursewares. For purposes of calculation, an SPSS macro was used kindly provided by J. D. Long (2007, personal communication). The results indicate that, across all included 153 participants between 2002/03 and 2004/05, for both coursewares "percentage of use" could play a more important role in predicting ranks than the mere hours of use (DOMR $w_{[M02\_082,\ \%]}$=.34, $w_{[K02\_022/23,\ \%]}$=.27, $w_{[K02\_014,\ hrs]}$=.01, $w_{[M02\_094,\ hrs]}$=-.04). However, with the overall associations found to be weak, the omnibus null hypothesis for the weights could not be rejected (Q^2=5.61<pF_{crit}9.73, p=.236).

In a second approach to the question posed above, measures of difference between the ex-post facto groups were calculated. Theoretically, if the notion that courseware use corresponds with superior situational judgment is correct, comparison groups with a higher degree of utilization should obtain higher situational judgement (SJT) post-test scores within this quasi-experimental setting.

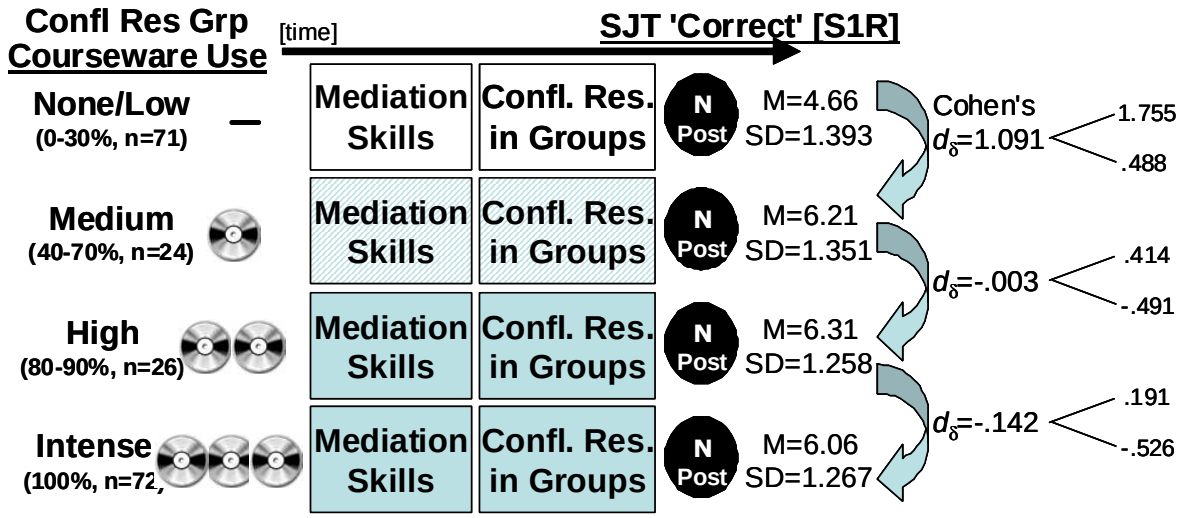


Figure 27. **Between ex-post facto group comparison (courseware use and situational judgement test sum score)**

The findings seem to suggest that this is generally the case; an example of comparison group x situational judgment post-test scores is depicted in Figure 27 for the ex-post facto courseware use groups (grouping alternative 'B' in Tables 26 and 27, see section 4.2) and one SJT sum score variable. Participant groups having used the courseware for learning purposes obtain significantly higher average scores in the post-test. Comparing the no-/low-use group to the other groups, it was found that the smallest proportion of group members to achieve higher

scores in the post-test than the no-/low-use group, minus the reverse, is $\delta_{[\text{None vs. Intense}]}$=.5186 or (x100%=) 51.86% (p<.001).

Interestingly, Figure 27 also seems to suggest that, between courseware-using participants, an increased use of the *Conflict Resolution in Groups* courseware may effect only marginal average gains – and a heavy usage may possibly entail detrimental effects in terms of SJT post-test scores. To enquire further, a second comparison was undertaken, this time with ex-post facto groups based upon the composite Courseware Usage Index CUI_B (see Table 25 in section 4.2) which additionally accounts for the use of the *Mediation Skills* courseware. Moreover, unlike the varying subgroup sizes in the previous analysis, the CUI_B ex-post facto groups are approximately equal in size. For purposes of comparison, the same variables as in Figure 27 were used.

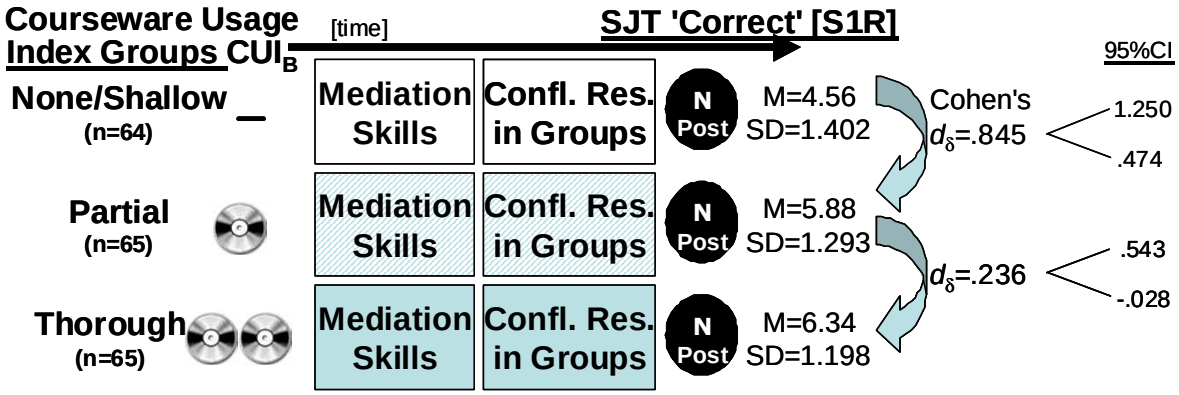


Figure 28. Between ex-post facto group comparison (situational judgement test sum score and composite Courseware Usage Index CUI$_B$)

The results shown in Figure 28 follow the same pattern as the ones presented previously. The proportional differences are markedly higher between the 'none' and 'partial' groups than between the 'partial' and 'thorough' groups; the effect size difference is significant ($\Delta\delta$=.3217, p=.005 two-tailed). There is an increase in average scores between the latter groups, nevertheless.

As the found post-test differences can also be explained by between-group differences uncontrolled for and possibly already present at pre-test, further cohort-based pre-post-analyses were undertaken for those cohorts where a situational judgement test was part of the pre-test. Due to the ordinal data level and the salient violations of normality assumptions, the traditional way of performing analyses of variance and/or covariance was barred as these are not robust enough. Alternatively, group differences were analysed by means of Cliff's (1996a, 1996b) ordinal delta (δ) statistic which, in effect, denotes the chance that test scores of sample X are

greater than the test scores of sample Y, minus the reverse. Thus, in effect, it is a measure of 'net' proportional superiority or directional dissimilarity (or non-overlap) of the distributions.

The following Figures 29 and 30 show the analyses for cohorts 2004/05 and 2005/06 median-split into ex-post facto groups according to self-reported courseware use (i.e. K_w in Table 25 in section 4.2). For informational purposes, means and standard deviations for the empirically scored situational judgement are also given.

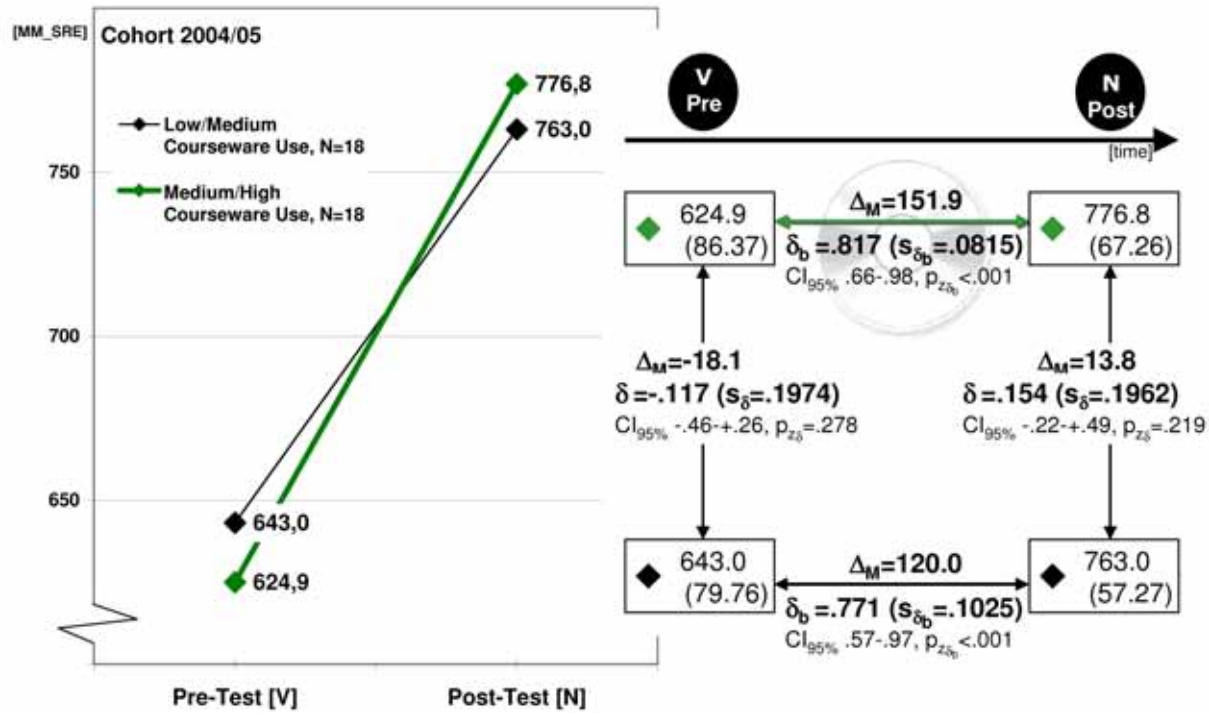


Figure 29. Pre-post analysis of SJT results for median-split courseware use ex-post facto groups 2004/05

For the 2004/05 cohort, no significant differences regarding situational judgement ([MM_SRE]) were found between the two groups neither at pre-test nor at post-test. (This also applies to all other situational judgement sum score variables.) As for within-group comparison, significant increases in SJT sum scores were found for both ex-post facto groups, however. Indeed, ordinal delta for the within-person change is $\delta_{w[K_w\ high]}$=100% for the "high courseware group", meaning that all participants in this group reached higher post-test scores when compared to their individual pre-test scores. In the "low" group, the same is true for only 83.3%, whereas 16.6% had even lower post-test scores resulting in $\delta_{w[K_w\ low]}$=(83.3-16.6=)66.7%. It has been argued (Cliff, 1996a, 1996b; Feng, 2007; Jeffrey D. Long et al., 2003; Romano et al., 2006) that the change of the groups as a whole, however, is possibly better expressed in terms of the distributional overlaps, or the (net) probability that a randomly selected post-test score is higher than a randomly selected pre-test score. These probabilities

are therefore detailed in Figure 29, with $\delta_{b[K_W\ high]}$=81.7% (p<0.01, Cohen's *d* with $CI_{95\%}$=1.309 < 2.033 < 4.071) versus $\delta_{b[K_W\ low]}$=77.1% (p<0.01, Cohen's *d* with $CI_{95\%}$=1.046 < 1.784 < 3.848).

In addition, the chart in Figure 29 seems to suggest that the mean increase is higher for the "high courseware use" group with $t_{c[K_W\ high]}(17)$=6.865 (p<.001) vs. $t_{c[K_W\ low]}(17)$=5.433 (p<.001 one-tailed). The difference between the probabilities reported above was not found to be significant, however, for the differences in group change ($\Delta\delta_b$=.817-.771=.046, p=.402, $CI_{95\%}$ -.316 to .408). In a direct between-group assessment of gain scores, 59% of the "high-usage" group attained higher scores than the "low-usage" group, compared to 41% for the inverse condition. (This analysis should possibly be treated with some care as it may be somewhat questionable if all changes in scores can be taken to mean the same). Again, no significant differences were detected between the groups (δ=58.9%-41.1%=.179, p=.364, $CI_{95\%}$ -.200 to .511).

In sum, the analysis did not yield results that support the stated hypothesis. However, it may be argued that, for the 2004/05 cohort, a K_W median split may not be a powerful enough measure to detect existing differences as self-reported courseware use in this cohort is both very high displays a low amount of variability: Of the 36 participants, 29 reported to having used the *Conflict Resolution in Groups* courseware to an extent of 80% or more, and of the remaining 7 students, 6 reported to have used the *Mediation skills* courseware to a degree of 70% or more.

In the 2005/06 cohort, self-reported courseware use variability (as also reported in Table 23 in section 4.2) is seemingly higher (usage of the *Mediation skills* courseware was not surveyed in 2006). In the sample base 2004/05 (n=36), $M_{2004/05}$=84.72 with SD=25.800 and Tukey's Q_{25}=85%, whereas in the cohort 2004/05 sample base (n=40) M=67.00 with SD=34.656 and Tukey's Q_{25}=35%, resulting in a significant F(1)=6.279 (p=.014). Moreover, the power of the SJT test used in 2005/06 was likely to be higher due to the revisions. Again, differences between the two groups at pre-test were tested for significance, again using the Mann-Whitney U test with exact p levels. Across all variables and pre-test sum scores, no significant differences were detected when using a Bonferroni-correction as suggested by Larzelere & Mulaik (1977). (Without any adjustment of α, the only significant difference found is a salient error sum scores in the Facilitation Knowledge subtest [V_Mod1F] with U=125.5, $|Z_U|$=2.047, p=.040). Neither were differences found for any of the situational judgment pre-test

scores, including the empirically scored [V_MM_SRE] depicted in the following Figure 30 (δ=-.010, p=.479) indicating a distributional overlap of approximately 99%.

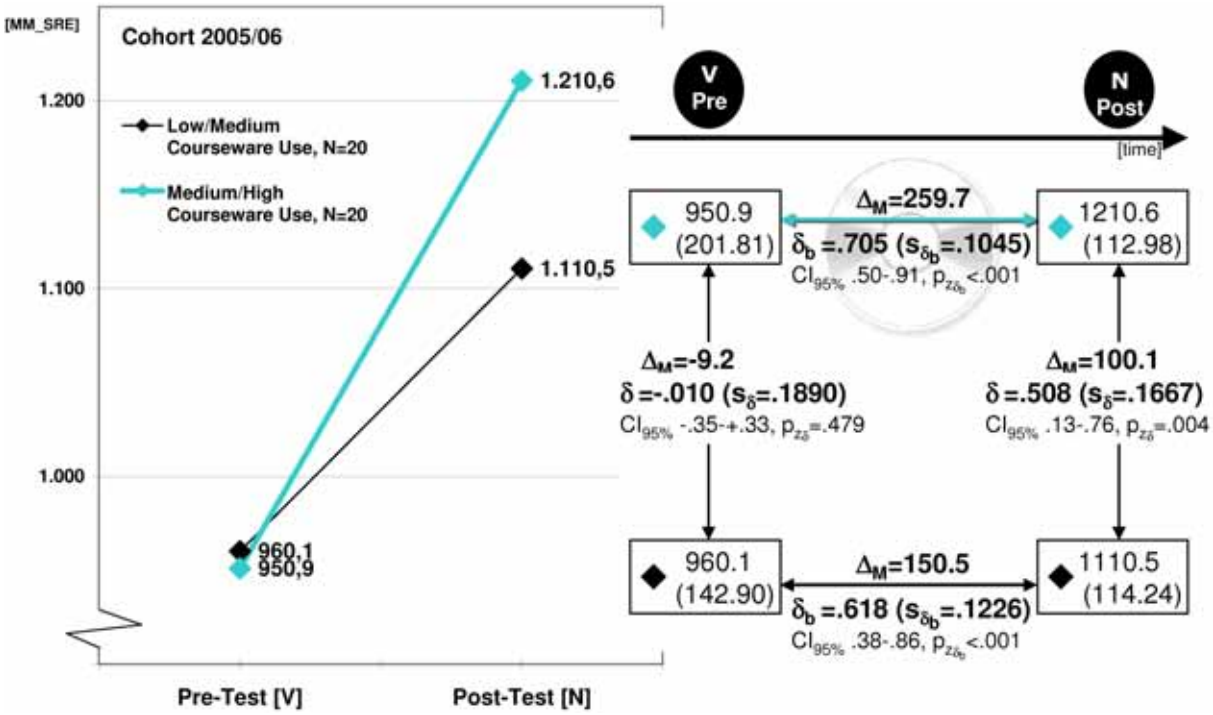


Figure 30. **Pre-Post analysis of situational judgement test results for median-split courseware use ex-post facto groups 2005/06.**

For both groups, significant increases between pre- and post-test were found as shown in Figure 30. However, at post-test, 75.3% of the high-usage group members arrive at scores superior to the ones of the low-usage group, compared to 24.5% of the scores being lower. This is a significant difference between the two groups (δ=75.3%-24.5%=.508, p=.004, Cohen's d_δ=.879 with $CI_{95\%}$=.177–1.710), expressed in more traditional terms as ΔM_{post}=100.1 and t(38)=2.785 (p=.004, $\tilde{\omega}^2$=.145, Cohen's d_t=.881). The difference in the extent of score gains was also found to be significant; approximately 2/3 of the "high-usage" group attained higher gains, compared to nearly 1/3 of the "low-usage" group (δ=67.3%-32.8%=.345, p=.037 one-tailed, with $CI_{95\%}$=.028–.600). Despite queries about non-normality, a repeated measures multivariate analysis of variance for repeated measures was performed to allow comparisons for the interested reader. The results reported in Table 30 below.

The MANOVA generally leads to the same conclusions, namely, that a significant main effect for learning is found, i.e. an increase between pre- and post-test scores ($F(1,38)=74.9$ with $p<.001$, on marginal means $t=8.655$, $p<.001$). Interestingly, based on mean estimates for the "low use" group of $M_{pre}=960.06$ / $M_{post}=1110.54$, and $M_{pre}=950.89$ / $M_{post}=1110.54$ for the "high use" group, respectively ($MSE_{pre}=39.098$ / $MSE_{post}=25.405$), the analysis also renders the interactive effect (Time/Learning x Courseware Use Group) significant with $F(1,38)=5.313$ ($p=.027$). However, the effect is small when compared to the main effect. Again, the between-subject comparison is not found significant as the difference between group mean estimates ($M_{low}=1035.30$, $M_{high}=1080.75$ with $MSE_{low/high}=28.395$) is a relatively small effect only ($F(1,38)=1.28$, $p=.265$, $\eta_p^2=.033$).

Table 30

Repeated Measures MANOVA for Pre-Post SJT Results and Ex-Post Facto Groups 2005/06

	Type III SS	df	MS	F	p	η_p^2
Time/Learning [V/N_MM_SRE]	841320.2	1	841320.20	74.910	.000	.663
Time/Learning x Courseware Use	59666.9	1	59666.89	5.313	.027	.123
Error (T/Learning)	426779.7	38	11231.04			

Notes. n=20 in both Courseware Use groups with total $M_{tl}=955.47$ ($SD_{tl}=172.659$) at pretest $M_{tl}=1160.57$ ($SD_{tl}=123.065$) at posttest. Equality of error variance across groups was assumed as Box's equality test was found insignificant for covariance matrices (Box'sM=6.74, p=.095) and Levene's test was neither found significant for pretest $F(1,38)= 3.042$ ($p=.089$) nor for posttest sum scores $F(1,38)=.071$ ($p=.792$). Sphericity was assumed for within-subjects effects. Within-subjects linear contrasts are equal to the within-subjects effects reported.

H2: Preparation and Situational Judgement.

Another question pertains to the combination of textbook and courseware preparation. As reported above, programme management expected the courseware, if used subsequently to textbooks, to serve illustrative and exemplifying purposes. Thus, one should expect to find superior situational judgement skills for users combining these preparational modes. In a related line of reasoning, higher outcomes should be associated with stronger use of both textbooks and coursewares.

In Table 31, an exemplary excerpt of the data and statistics calculated is presented to highlight the prevailing empirical situation in relation to this issue. Table 31 also details multiple regression coefficients (based on Spearman's rank correlations) for both the textbook/literature and courseware usage indices. The only significant multiple R found is the one in the 2005/06 cohort ($R^2_{adj}=.231$, $p=.008$). Here, the courseware predictor coefficient was also found significant ($\beta=.443$, $t=3.082$, $p=.004$), in contrast to the [K_Lit]-textbook

preparation index (β=-.230, t=-1.590, p=.120). Interestingly, however, while the predicted situational judgement variable seems to be positively associated with courseware use ($r_{s_{N\_M\_S1RxK\_Lit}}$=.422) , the extent of textbook or literature use is negatively associated with post-test scores ($r_{s_{N\_M\_S1RxK\_Lit}}$=-.190; y[N_MM_S1R]=5.339–.194[K_Lit]+.031[CUI_K]). This finding is somewhat in contrast to the – albeit not significant – findings for the other cohorts and variables.

Table 31

Prediction of Post-test Situational Judgement Scores by Preparation Indices for Sample Base (Multiple Regression Analysis)

	Predicted SJT post-test variable			β (Standardized Multiple Regression Coefficients)		ANOVA			
Cohort	**[N]**	**R**	**Adjusted R^2**	**Textbooks [K_Lit]**	**Courseware [CUI_K]**	**Regression Sum/Sq.**	**Mean Sq.**	**F[a]**	**p**
2002/03	MM_S1R	.137	-.001	.137	—	1.832	1.832	.963	.331
2003/04	MM_S1R	.192	.006	-.189	.119	3.412	1.706	1.187	.312
2004/05	MM_S1R	.247	.004	.176	-.197	2.776	1.388	1.069	.355
2005/06	MM_S1R	.481	.231	-.230	.443	21.799	1.899	5.555	.008
2002/03	MM_S2R	.025	-.019	-.025	—	.114	.114	.032	.859
2003/04	MM_S2R	.252	.033	-.246	.163	11.818	5.909	2.100	.130
2004/05	MM_S2R	.234	-.003	.136	.174	7.125	3.562	.956	.395

Included number of participants were n=52 for 2002/03, n=65 for 2003/04, n=36 for 2004/05, and n=40 for the 2005/06 cohort. A revised version of the post-test had been used in 2005/06. [a]df=1 in 2002/03 and df=2 in the blended learning cohorts.

Another approach to the issue under discussion is a comparison of the test scores of the ex-post facto groups based on both the textbook and literature preparation (see Table 28 in section 4.2) and the courseware use indices (see Table 25 in section 4.2). By use of the between indices, 3 groups (low-medium-high) can be distinguished for both preparational modes; resulting in 9 groups in a 3x3-table (see Figure 31 below). Between-group comparisons were calculated for situational judgment scores MM_S1R, MM_S2R and MM_SRE and cohorts 2002/03 through 2004/05 and the revised post-test version 2005/06. The analyses generally follow the same pattern exemplarily depicted in Figure 31 for MM_S2R and the sample base cohorts 2002/03 through 2004/05. Three main conclusions can be drawn based on the data available.

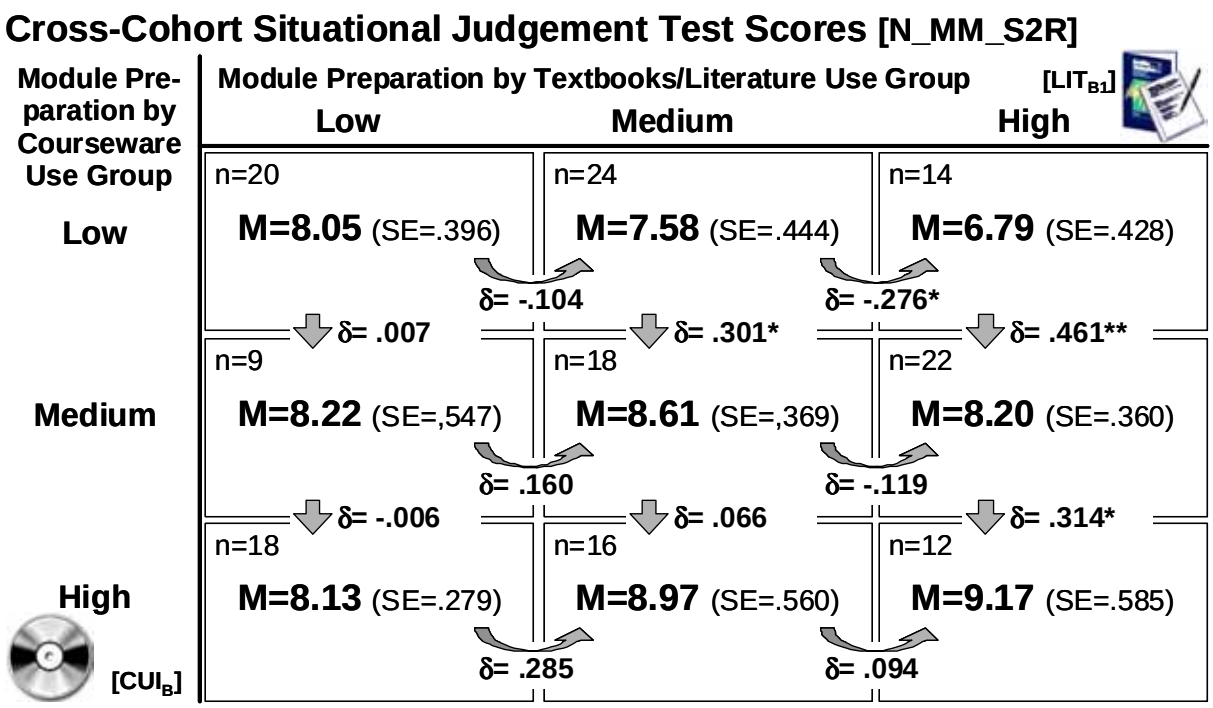


Figure 31. **Situational judgement test scores for Ex-post Facto Groups based on Courseware and Textbook Use. *p<.05, **p<.01. Based on sample base cohorts 2002/03 through 2004/05 (N=153).**

First, ex-post facto groups with at least medium courseware use display improved situational judgement as compared to all three 'low-courseware use' groups. This effect was repeatedly and consistently found across all cohorts and for all situational judgment post-test scores listed above. Averaging the effect sizes (and their respective 95% CIs) across all low-to-medium differences (as measured by Cliff's ordinal dominance δ) and the three outcome variables for sample base cohorts 2002/03 through 2004/05 yields a mean Cohen's d_{δ}=.593 (with a $CI_{95\%}$ ranging from -.014 to 1.352).

Second, CSSL courseware should be used as a complementary, not substitutional preparational means. No significant differences were found between the courseware use groups that reported "low" use of reading materials recommended by programme management.

And, third, intensified cramming by use of written materials may impede situational judgement: High-textbook-use groups tend to achieve lower scores than medium-use groups unless courseware use is not high. Again, this effect was repeatedly and consistently found across all post-test versions and for all situational judgment post-test scores. Here, the average effect size expressed as Cohen's d_{δ} is =-.248 (with a $CI_{95\%}$ ranging from -.546 to .263).

H3: CSSL and Knowledge.

Does courseware use correspond with more detailed (declarative) knowledge of the concepts and the taxonomy of the mediation model presented in the curriculum? And is it associated with a higher level of interest in the subject of mediation and conflict resolution at the end of the curriculum? Using the same approaches to analysis as above, an examination of data in regard to these questions is undertaken in the following.

The main post-subtest for mediation and conflict resolution process knowledge was N-6 as found in the Appendix C (Sec II, p. 12) the first two part of which as summarized in sum scores [N_6] and [N_6corr] focused on knowledge detailed in various books on the subject (Besemer, 1997; R. Fisher, Ury, & Patton, 1991; R. Fisher, Ury, & Patton, 1998; Glasl, 1999; Redlich, 2004a; Thomann, 1998; Thomann & Schulz von Thun, 1988). Basic knowledge thought to be concomitantly imparted through use of the courseware was nevertheless attempted to be tapped in the post-tests, beginning in 2004/05 and represented by sum score [N_6_2]. (The 'total' score [N_6_3] represents the combination of both sum scores.)

Spearman correlation coefficients and effect size estimates for the association between these mediation knowledge sum scores and both courseware and textbook use indices are detailed in Table 32. Across cohorts and in most individual cohorts, use of the mediation-related coursewares (CUI_{TTL}) is positively associated with mediation knowledge post-test scores with r_s ranging between .160 and .425 for those r_s significantly differing from zero. Expressed in standardized effect size units, Cohen's d_{r_s} range between .325 and .938. Puzzling, however, is one significant negative association found in the 2004/05 cohort (r_s=-.288).

Similar to the analyses conducted to investigate the relationship between situational judgement and the various preparation modes (H2), a dominance-based ordinal multiple regression (DOMR) was also calculated for mediation knowledge, using courseware use and self-reported degree of reading as predictors. More specifically, the ranked mediation knowledge total test score N_6_3 was predicted from the mediation-related courseware use index (CUI_{TTL}) and the degree of reading [K04b_02] for cohorts 2004/05 and 2005/06.

Table 32

Association (r_s) between Use of Learning Materials and Mediation Knowledge Post-Test Scores

Sample Base Sum Score	Redlich (1997) Ch3 [M02_03]	Redlich (1997) Ch1/2 [K04b_02]	CUI_K	CUI_M	CUI_{TTL}	CUI_{TTL} Effect Size Estimate[a] Cohen's d_{rs}	95% CI d_{rs} lower	95% CI d_{rs} upper
2003/04 (n=66)								
N_6	-.003	.013	.175	-	.175	.354	-.156	.888
N_6resp	-.009	-.085	.018	-	.018	.035	-.478	.550
N_6corr	-.026	.103	.211 *	-	.211 *	.430	-.082	.970
2004/05 (n=36)								
N_6	.095	-.211	-.097	.042	.006	.011	-.705	.729
N_6resp	-.017	-.171	-.024	.218	.230	.472	-.235	1.238
N_6corr	.170	-.149	-.196	-.189	-.288 *	-.601	-1.387	.110
N_6_2	.036	-.158	.068	-.231	.086	.172	-.537	.902
N_6_3	.152	-.215	-.056	-.252	-.064	-.128	-.855	.582
2002/03-2004/05 (n=155)								
N_6	-.003	-.049	.148 *	.115	.162 *	.329	.000	.667
N_6resp	-.061	-.089	-.065	.111	-.003	-.007	-.336	.322
N_6corr	.030	.006	.215 *	.060	.193 *	.393	.064	.734
2005/06 (n=40)								
N_6	-	-.031	.288 *	-	.288 *	.601	-.071	1.340
N_6resp	-	-.028	.185	-	.185	.377	-.290	1.085
N_6corr	-	-.007	.198	-	.198	.404	-.263	1.116
N_6_2	-	-.164	.045	-	.045	.091	-.581	.772
N_6_3	-	-.141	.105	-	.105	.211	-.457	.902
2002/03-2005/06 (n=195)								
N_6	-.003	-.136	.085	.166 *	.132 *	.266	-.026	.563
N_6resp	-.061	-.143	-.079	.144 *	-.009	-.018	-.310	.274
N_6corr	.030	-.063	.147 *	.106	.160 *	.325	.032	.624
N_6_2[b]	.036	-.306 *	.103	.420 *	.425 *	.938	.438	1.491
N_6_3[b]	.152	-.316 *	.105	.379 *	.385 *	.836	.342	1.376

* p < .05; ** p < .01; *** p < .001 (one-tailed). For sample base cohort 2002/03 (no blended learning), CUI=0 was assumed. [a]Effect size and confidence interval estimates calculated as $d_s = 2r_s(1-r_s^2)^{-1/2}$. The CI values are based on Fisher-Z-transformation for Spearman's r_s as detailed in Zar (1999; cit. in Sheskin, 2007, p. 1364) with $CI_{Z_{s(95\%)}} = 0.5 \ln[(1+r_s)/(1-r_s)] \pm 1.96\,[1.060/(n-3)]^{1/2}$. [b]n=76 for 2004/05 and 2005/06 only.

The DOMR findings indicate that use of reading materials may be a slightly more important predictor than degree of courseware use with $w_{[K04b\_02]}$=.22 compared to $w_{CUI_{TTL}}$=.18, with rejection of the omnibus null hypothesis for the weights (Q^2=9.050>pFcrit 6.234, p<.001). For purposes of comparison, a "traditional" least-squares multiple regression analysis was calculated despite the discreteness of the variables entered, the ordinal level of measurement and non-normality reservations. Entering the same variables as above yields R=.430 (adjusted R^2=.162 with F(2,73)=8.275 and p<.001). Interestingly, in contrast to ordinal regression, the analysis suggests courseware use to be a more influential predictor (β=.311) than preparatory reading (β=-.223).

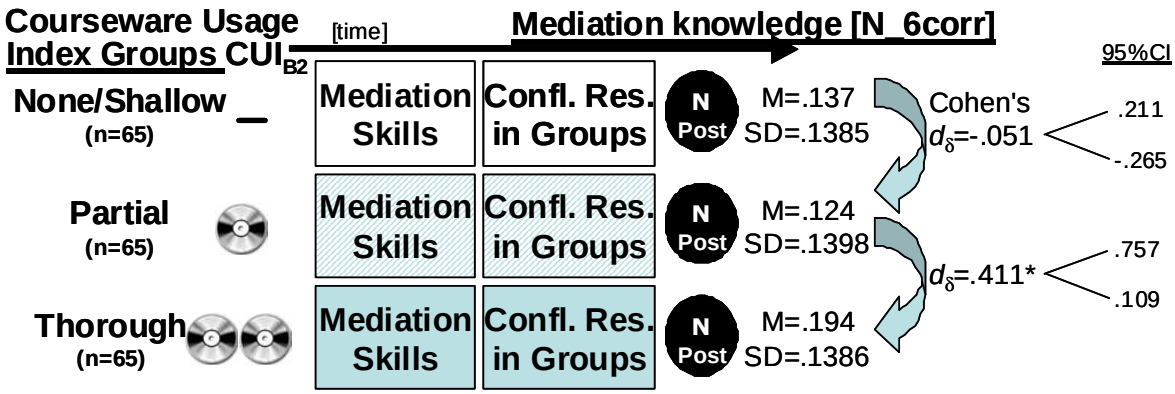


Figure 32. **Ex-post facto group comparison (courseware use and chance-corrected mediation knowledge test sum score)**

A significant difference was also found between the post-test scores attained by the "high courseware use group" and the other two groups based on tercile-split. This applies to both ordinal dominance analysis ($\delta_{hi/med}$=.280, $p_{hi/med}$=.002, Cohen's d_δ=.411; $\delta_{hi/low}$=.240, $p_{hi/low}$=.007, Cohen's d_δ=.343) as well as the "traditional" t-test statistic with $t_{hi/med}(128)$=2.885 ($p_{hi/med}$=.002, Cohen's d_t=.506) and $t_{hi/low}(128)$=2.397, $p_{hi/low}$=.009, Cohen's d_t=.420). No significant mediation knowledge post-test score differences were found between the "low" and "partial use" groups (see Figure 32).

H4: CSSL and Post-Curricular Level of Interest.

The post-tests instruments also included questions pertaining to the curiosity the respective participant had developed in the subject matter. Participants were asked to rate their (current) level of personal interest they take in the range of topics addressed throughout the *PROCON* curriculum on a 6-point Likert scale (see Appendix C [N], Sec II-4, p. 5; [N_401…05]). Low values indicate a strong lively interest, high values the reverse. In case the level of interest is associated with courseware use, one should expect to find empirical sings of this assumed relationship such as negative correlations between courseware use variables and the Likert scale scores.

Ideally, the Likert scales would have been part of the pre-tests. However, thorough discussions with programme management and peer tutors had given rise to the concern that pre-test answers to these scales would likely be subject to considerable social desirability bias. It was also argued that, as the *PROCON* curriculum was an add-on to the usual course of study, students applying for participation would be those most interested in the subject anyway. It was therefore decided to exclude the items from pre-test instruments. Following the spirit of the argument that interest in participation may have been highest in those displaying intense

preparation for the pre-tests, one alternative indicator would be the number of hours spent for pre-test preparation ([V_24oMR], see Appendix A, Sec I, p. 3, Item 24). Thus, partial rank correlation coefficients were calculated between self-reported use of the *Conflict Resolution in Groups* courseware ([K02_022/23], [K02_014], CUI_{TTL}) and the post-test level of interest in the various aspects of mediation theory and practice ([N_402…05]). The results are summarized in Table 33.

Table 33

Association between Courseware Use and Post-test Interest (Controlling for Pre-test Preparation)

Cohort and Courseware Use Variable	Interest in… …Subject 'Mediation' [K_402]	…Theories Confl. Res. [K_403]	…Confl. Res. Practice [K_404]	Future Preoccupation [K_405]	Z-averaged Partial r_s	Average Effect Size[a] Cohen's d_{rs}	95% CI d_{rs} lower	upper
2003/04 (n=60)								
Percent Used[b]	-.184	-.018	-.001	-.168	-.094	.188	-.349	.738
Hours Used[c]	-.081	-.185	-.043	-.142	-.113	.228	-.308	.781
CUI_{TTL}	-.063	-.104	.025	-.156	-.075	.150	-.387	.698
2004/05 (n=29)								
Percent Used[b]	-.335 *	-.172	-.181	-.314 *	-.252	.520	-.278	1.401
Hours Used[c]	-.203	-.040	-.089	-.221	-.139	.281	-.517	1.124
CUI_{TTL}	-.272	-.070	-.105	-.246	-.174	.354	-.443	1.207
2005/06 (n=37)								
Percent Used[b]	.067	.130	.171	.244	.154	-.311	-1.056	.395
Hours Used[c]	.006	.199	.350 *	.159	.181	-.369	-1.121	.337
CUI_{TTL}	.042	.185	.315 *	.207	.189	-.385	-1.139	.322
All Blended Learning (n=142)								
Percent Used[b]	-.189 *	-.080	-.047	-.091	-.102	.205	-.139	.555
Hours Used[c]	-.229	-.171 *	-.048	-.127	-.144 *	.292	-.053	.645
CUI_{TTL}	-.177 *	-.159 *	-.053	-.097	-.122	.245	-.099	.596

* p < .05; ** p < .01; *** p < .001 (one-tailed). Coefficients are partial correlation coefficients calculated from Spearman's r, controlling for the number of hours spent pre-test preparation [V_24oMR]. [a]Average effect size (d_{rs}) and confidence interval estimates were calculated as in Table 30 based on negative Fisher-Z-averaged Spearman's r across items [N_402…05]. Higher values indicate higher degrees of association. [b]Conflict Resolution in Groups Online Module Survey variable K02_022/23 (see App. B). [c]Conflict Resolution in Groups Online Module Survey variable K02_014 (see App. B).

Controlling for pre-test preparation, there is a significant negative average relationship found between the number of hours and the post-test level of interest items across all blended learning cohorts (mean r_{sp}=-.144, p=.043, Cohen's d_{rs}=.292). However, contrary to expectation, the associations were found to be positive throughout in the 2005/06 cohort. This finding is sustained by between-group analyses. As depicted in the following Figure 33, there may be a systematic (albeit not always significant) deviation in the 2005/06 cohort from the expected superiority in post-test levels of interest of the high courseware groups.

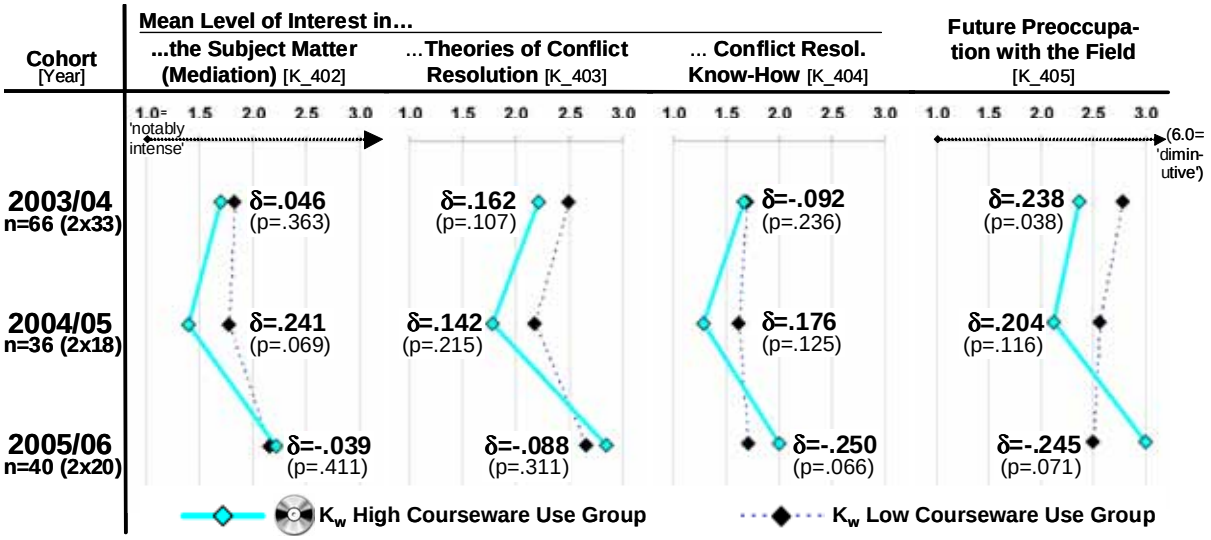


Figure 33. Post-test level of interest in subject matter: Comparison between ex-post facto courseware use groups.

H5: CSSL and Learning Gains.

In all module online surveys, participants were asked to report and rate their learning gains in areas thought likely or possibly to be addressed by the respective module. Results for the various modules were reported in section 4.1. In the following, the influence of preparatory use of the *Conflict Resolution in Groups* courseware on self-reported learning gains in the respective module is investigated.

The first analysis seeks to estimate the cross-cohort association between learning gains and various variables thought to predict the amount of gains. It has been an additional objective of this approach to evaluate the relative importance of courseware use in comparison to the extent of preparation by textbooks/literature, and the quality of the subsequent classroom training sessions. Therefore, despite the probable violations of test assumptions (given the ordinal level of measurement on the predicted gain score variables and potential non-normality and heteroscedasticity problems), least-squares multiple regression analyses were calculated. All cohorts were included in these analyses, i.e. with CUI_K=0 for the traditionally delivered (or non-blended-learning) cohort 2002/03. The results detailed in Table 34 below can be summarized as follows:

- Generally, weak significant associations are found between gain scores and the combination of the three predictor variables (maximum R^2_{adj} found =.167 in [K09_01, "guiding a group through the stages of the mediation process"]).
- Relative to those of the preparation scores predictors, the standardized regression weights (β values) are notedly higher for the overall classroom-based training score variable.
- Across all gain score items and cohorts, no regression coefficients significantly differing from zero were found for the courseware use predictor variable.

- With few exceptions, the standardized regression weights of the courseware use predictor variable are generally lower than those of the textbook preparation predictor variable.

Table 34

Cross-Cohort Prediction of Learning Gains in the* Conflict Resolution in Groups *Module (Multiple Regression Analysis)

						β (Standardized Multiple Regression Coefficients)		
Predicted variable	**Gain[a] M (SD)**	**R**	**Adjusted R^2**	**ANOVA F^b**	**p**	**Courseware Use [CUI_K]**	**Textbook Use [K_Lit]**	**Classroom [K14][c]**
K09_01	2.6 (.90)	.422	.167	16.372	<.001	-.110	-.112	-.389 ***
K09_02	3.1 (1.19)	.037	.001	.101	.959	.009	.027	-.024
K09_03	2.5 (.95)	.255	.053	5.282	.002	-.080	-.038	-.236 ***
K09_04	2.4 (.98)	.231	.041	4.266	.006	-.079	-.065	-.205 **
K09_05	2.3 (.88)	.268	.059	5.850	.001	-.087	-.166	-.195 *
K09_06	2.6 (.98)	.150	.023	1.747	.158	-.038	-.094	-.113
K09_07	2.7 (1.14)	.296	.076	7.285	<.001	-.039	-.125	-.267 ***
K09_08	2.6 (1.02)	.279	.066	6.377	<.001	-.032	-.075	-.267 ***
K09_09	2.9 (1.11)	.338	.103	9.765	<.001	.011	-.135 *	-.312 ***
K09_10	2.9 (1.11)	.223	.037	3.943	.009	.008	-.078	-.210 **
K09_11	3.3 (1.23)	.135	.005	1.396	.245	-.078	-.063	-.090
K09_12	2.7 (1.15)	.258	.054	5.394	.001	-.032	-.081	-.243 ***
K09_13	2.4 (.91)	.316	.088	8.381	<.001	-.108	-.170 **	-.246 ***
K09_14	2.8 (1.07)	.183	.021	2.614	.052	-.084	-.066	-.148 *
K09_15	2.5 (1.13)	.279	.066	6.403	<.001	-.103	-.063	-.249 ***
K09_16	3.0 (1.14)	.214	.033	3.626	.014	-.011	-.137 *	-.166 *
K09_17[d]	2.4 (.95)	.224	.032	2.823	.041	.039	-.149	-.165 *
K09_18[d]	2.5 (.92)	.265	.053	4.026	009	.086	-.137 *	-.212 **

Notes. *p<.05, **p<.01, ***p<.001 (two-tailed) for coefficient t-tests with df=226 (df=160 for items K09_17 and K09_18). Included number of participants were n=66 for 2002/03, n=77 for 2003/04, n=42 for 2004/05, and n=42 for the 2005/06 cohort.

[a]Progress as reported on a 6-point Likert scale ranging from 1-("very high") to 6-("very low").

[b]F for regression/R^2 with *df*=(3,227). For items K09_17 and K09_18, *df*=(3,160).

[c]Classroom Training overall score as rated by participants on a 15-point grading scale used in German schools with 15 points ("1+") denoting best possible achievement.

[d]Blended learning cohorts (2003/04, 2004/05, 2005/06) only.

Another variable with potentially predictive value is the self-reported level of experience prior to module participation [K_01]. In all blended-learning cohorts, participants were asked to select their prior level of understanding from a drop-down list (see Appendix B, [K] for details). To these selections, scores between 1 and 6 were assigned. For example, the selection of "I have no knowledge and no experience with group-based conflict resolution" was assigned a score of 1 and "I am regularly/professionally assuming the role of team conflict manager/facilitator" was assigned a score of 6. However, analyses reveal that entering this variable into the regression equation leads to only marginal increases in R^2_{adj}. Significant positive β–weights for this predictor were found in item [K9_02, "conduct contracting adequately"] (β=.177, p=.027), [K9_05, "transparent structuring of the process"] (β=.155, p=.046) and [K9_11, "safeguarding solution implementation"] (β=.163, p=.041).

In a second approach to investigating the influence of courseware use on learning gains, differences between low, medium, and high courseware use groups were analysed. For this purpose, by use of courseware use index K_{B2} (see Table 25 in section 4.2), all cohorts were included in the analysis. Gain scores averages for the tercile-split ex-post facto groups (i.e. "low", "medium", "high" courseware use) as well as between-group differences as expressed in Cliff's δ are depicted in Figure 34 below.

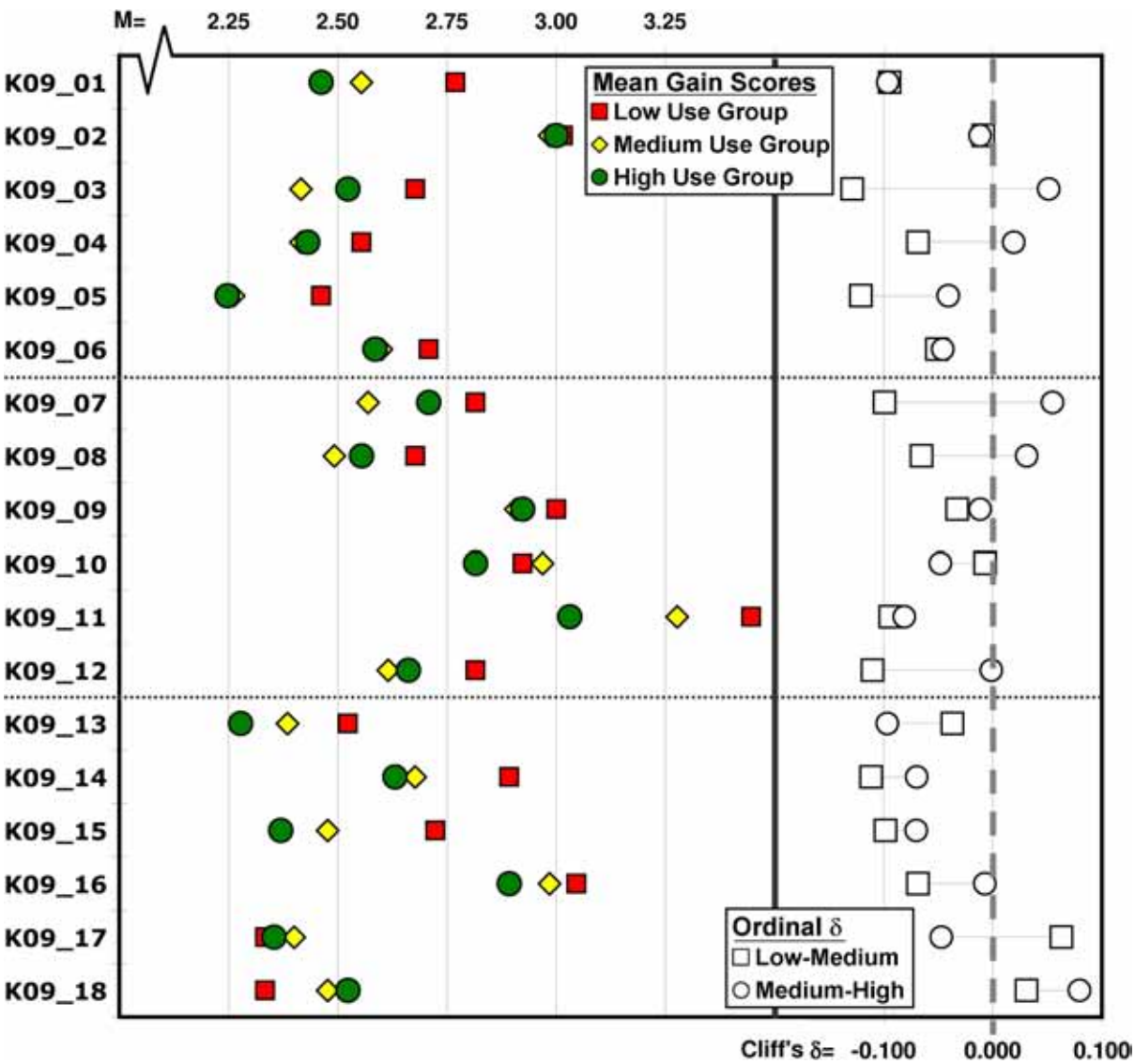


Figure 34. Cross-blended learning cohorts learning gain score averages (K09) and ordinal dominance statistics for between ex-post facto courseware use groups (tercile-split/K_B).

For items K09_01 through K09_16, the lowest average gain scores are found in the "low use" ex-post facto group. For these items, ordinal dominance statistics reveal that, without exception, a majority of "low" group members reports lower gains than the members of both other groups. Significant ordinal score gain differences between the "low" use group and the two other groups were found in items [K09_01, "guiding a group through the stages of

the mediation process"] ($\delta_{low/med+high}$=-.140, p=.046), [K09_05, "transparent structuring of the process"] ($\delta_{low/med+high}$=–.139, p=.039), [K09_11, "safeguarding solution implementation"] ($\delta_{low/med+high}$=–.138, p=.049), [K09_14, "concise and fluent visual recording"], and [K09_15, "maintaining a sense of neutrality"] ($\delta_{low/med+high}$=–.131, p=.060). Across all items, this effect for the "low" vs. "med/high" comparison is $\delta_{low/med+high}$=–.071 or, as expressed in Cohen's d_{δ}=.086.

For two additional items, the order of the means and the d values depicted in Figure 34 are consistent with the hypothesis that courseware use may have a bearing on learning gains. These are [K09_13, "structuring of the entire mediation process"] ($\delta_{low/med}$=–.037, p=.347; $\delta_{med/high}$=–.097, p=.153) and [K09_16, "steering others towards more accordant attitudes"] ($\delta_{low/med}$=–.069, p=.242; $\delta_{med/high}$=–.008, p=.469).

Ideally, these findings that "low" courseware use group gains tend to be lower than those of the "medium" and/or "high" courseware ex-post facto groups, are replicated in analyses by single-blended-learning-cohorts. However, using K_w median split ex-post facto groups "low" vs. "high", the results are found consistent for only two items ([K09_01] and [K09_15]) within the blended learning cohorts 2003/04 ($\delta_{[K09\_01]}$=-.054, p=.342; $\delta_{[K09\_15]}$=-.073, p=.295), 2004/05 ($\delta_{[K09\_01]}$=-.056, p=.380; $\delta_{[K09\_15]}$ =-.117, p=.268), and 2005/06 ($\delta_{[K09\_01]}$=-.190, p=.127; $\delta_{[K09\_15]}$ =-.173, p=.164).

In addition, another between-group difference found consistent both within and across cohorts is the seemingly higher average gain for the "low" courseware group in item [K09_18, "counselling for individuals and groups in general"] ($\delta_{low/med+high}$=+.090 or, as expressed in Cohen's d_{δ}=.118). However, items [K09_17] and [K09_18] were introduced for purposes of control only, and these were not considered direct learning objectives of the *Conflict Resolution in Groups* modules.

DIGEST 4.3

The additional use of the mediation-related CSSL coursewares in the respective training modules was expected to effect superior performance levels in post-curricular situational judgement tests (SJT). Significant associations were found between courseware use and SJT 'pick-best' scores ($r \approx +.4$ across cohorts). The highest post-test differences were found between 'low' and 'medium' (or 'partial') use groups (Cohen's $d \approx +.8$ to +1). In 2005/06, a revised and potentially more powerful SJT version was employed in pre- and post-tests; here, significant differences between courseware use groups were found for both pre-to-post score gains (Cohen's $d \approx +.5$) and the post-test SJT score level ($d \approx +.8$).

It was also expected that the combined use of CSSL and traditional preparational materials effects improved situational judgement. The overall findings lend support to this assumption. Again, significant SJT post-test score differences were found between "low" and "medium" use groups (Cohen's $d \approx +.5$), but only where CSSL courseware was used as a complementary, not substitutional means of preparation (i.e. "medium" or "high" textbook use). In addition, the "high" literature use groups displayed significantly lower SJT scores than "medium" groups unless CSSL courseware use was also "high" (Cohen's $d \approx -.2$).

Another expected impact of CSSL courseware use was improved knowledge of the concepts and the taxonomy of the mediation model presented in the curriculum. The findings generally lend support to the notion that courseware use may have an impact, but it was found weak, too indefinite and possibly restricted to 'thorough' users of the coursewares. Similarly, the post-curricular level of interest in the subject matter (i.e. mediation, conflict resolution) was not consistently found higher for "high" CSSL courseware use groups.

4.4 CSSL and Self-Efficacy

H6: CSSL and Self-Efficacy Levels.

As detailed above, significant mean increases were reported by participant groups which, across cohorts, neatly correspond to the subject matter imparted in one or several *PROCON* modules. However, it was pondered about the roles computer-supported learning might play within this context. On the one hand, those participants with high CSSL usage may report higher increases in self-efficacy possibly due to a more intense module preparation and the preliminary exposure to and the training foci on behavioural models presented. On the other hand, these participants may be equipped with a higher level of self-knowledge or an awareness as to discrepancies between the (most effective) behavioural models shown and their own perceived skills. This, in turn, might lower the self-rated confidence in one's own abilities. If, either way, there is a significant influence of courseware use, one should expect to find associations between courseware use and changes in self-reported self-efficacy level. By contrast, the vast majority of coefficients reported in Table 35 have not been found significant. Neither do the signs of the found correlations provide helpful clues which could help in answering the question posed above.

Table 35

Association between Courseware Usage and Self Efficacy Raw Gains

SE-Item/Content Description	PCOI[c]	2003/04[a]				2004/05[b]			
		CUI_{TTL}		Alt B Group		CUI_{TTL}		Alt B Group	
		r_s	p	r_s	p	r_s	p	r_s	p
01 Counselling/Interpers. Knowledge	V-N	-.076	.544	-.114	.361	-.256	.132	-.344	.040
02 Applied Counselling/Interpers. Skills	V-N	-.086	.491	-.081	.516	-.099	.564	.041	.812
07 Mediation Competencies Knowledge	P-M	.019	.881	-.111	.376	-.069	.689	.025	.885
08 Conflict Analysis Knowledge	P-N	-.078	.533	-.192	.123	-.141	.412	-.127	.459
09 Conflict Management Knowledge	P-N	-.004	.974	-.028	.822	-.106	.539	-.065	.707
10 Mediation Strategy Knowledge	P-M	.070	.575	.019	.877	.008	.961	.049	.779
11 Conflict Moderation Knowledge	P-K	.033	.793	-.058	.646	.018	.915	.104	.546
12 Mediation Skills	P-M	.104	.405	.032	.798	-.213	.211	-.140	.416
13 Conflict Resolution in Groups Skills	M-K	.074	.554	.018	.888	.097	.574	.071	.679

Spearman rank correlation coefficients with two-tailed probability levels for association between raw score gains (on 6-point Likert scale with 1-"My practical skills are notedly comprehensive" … 6-" My practical skills are rather limited"; gains were calculated as "pre"-"post" for gains to be expressed as positive values) and either CUI_{TTL} courseware usage index or affiliation with courseware group level (see Table 26, Alt. B, with 1-low…4-high usage). [a]n=66. [b]n=36. [c]Paired comparison of interest, i.e. stating post-module surveys between which the raw score gain had been calculated. The PCOI corresponds to the subject matter of certain *PROCON* modules (with V-Pre-test, F-Facilitation module online survey, P-Problem Solving module online survey, M-Mediation Skills module online survey, K-Conflict Resolution in Groups module online survey, N-Post-test).

A similar inconclusive picture emerges when the 2005/06 cohort is included in the analysis. Associations between Pre- to Post-test gains and courseware use are reported in the following Table 36.

Table 36

Association between Courseware Usage and Pre- to Post-test Raw Gains in Self Efficacy

SE-Item/Content Description	2003/04[a]	2004/05[b]	2005/06[c]
01 Counselling/Interpers. Knowledge	-.076	-.256	.031
02 Applied Counselling/Interpers. Skills	-.086	-.099	-.070
07 Mediation Competencies Knowledge	-.136 [d]	.018	.278
08 Conflict Analysis Knowledge	-.247 *	-.121	.304
09 Conflict Management Knowledge	-.221	-.082	.074
10 Mediation Strategy Knowledge	-.036 [d]	.018	.329 *
11 Conflict Moderation Knowledge	.033 [d]	.018 [d]	.112
12 Mediation Skills	.016	.082	.221
13 Conflict Resolution in Groups Skills	-.238	.042	.165

Spearman rank correlation coefficients with two-tailed probability levels for association between CUI™ courseware usage index and pre- to post-test gains in raw scores (on 6-point Likert scale with 1-"My practical skills are notedly comprehensive" … 6-" My practical skills are rather limited"; gains were calculated as "pre"-"post" for gains to be expressed as positive values). *p<.05. [a]n=66. [b]n=36. [c]n=40. [d]Gains calculated from *Problem Solving* and *Conflict Resolution in Groups* Module Surveys.

Tightening the working hypothesis, usage of the coursewares should affect those self-efficacy levels most closely related in terms of the content. For the *Conflict Resolution in Groups* courseware these are SE-items 10 (Mediation Strategy Knowledge), 11 (Conflict Moderation Knowledge), and 13 (Conflict Resolution in Groups Skills). As for the *Mediation Skills* courseware employed in the 2004/05 blended learning module, these are SE-items 7 (Mediation Competencies), 10 (Mediation Strategy Knowledge), and 12 (Two-Party Mediation Skills). In the same cohort, the *Problem Solving* courseware was used. Items pertaining to this *PROCON* module are SE-items 05 (Problem Solving Knowledge) and 06 (Applied Problem Solving Facilitation Skills).

For these items, further cohort-based pre-post-analyses were undertaken for the respective blended learning cohorts. Again, salient violations of normality assumptions could not be ruled out, and the ordinal data level of measurement admits analysing group differences by means of Dominance Analysis. The following Figures 35 through 38 depict the outcome of the cohort-wise analyses for median-split ex-post facto courseware usage groups for the *Conflict Resolution in Groups* courseware (i.e. K_w in Table 25 in section 4.2). In 2003/04 and 2004/05, the self-efficacy items were part of the online surveys. Participants in these cohorts were asked to fill these in after the *Problem Solving* Module [P] (i.e. prior to the *Mediation Skills* Module) and after the *Conflict Resolution in Groups* Module [K]. In 2005/06, self-

efficacy items were only implemented in the pre- [V] and post-tests [N]. Again, SE-item means are also detailed in the following Figures.

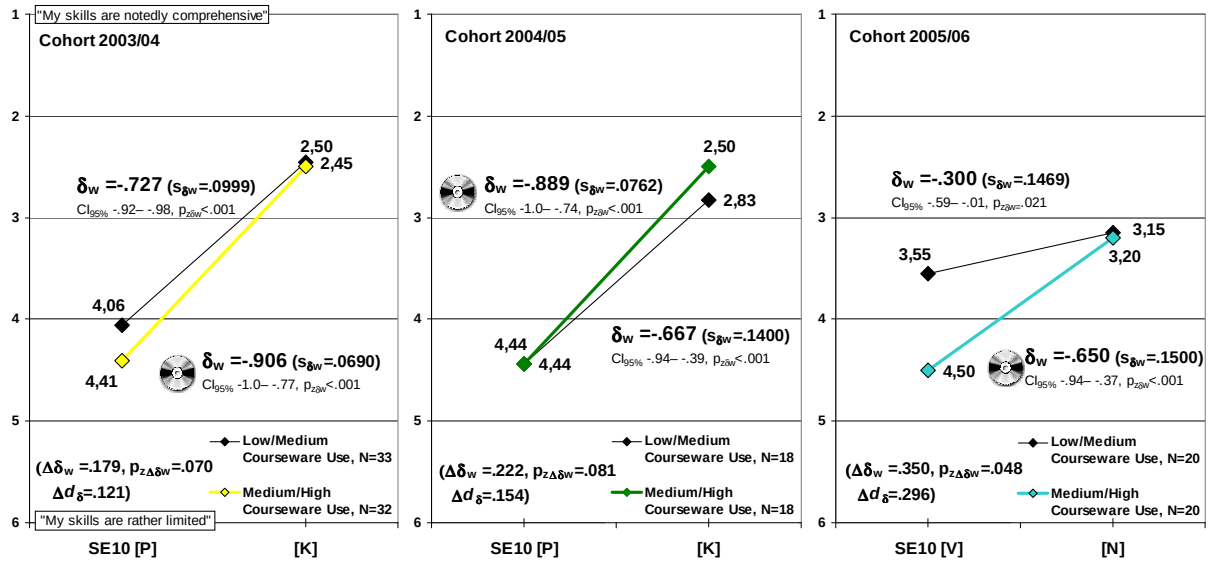


Figure 35. **Changes in mediation strategy self-efficacy levels [SE10] for median-split ex-post facto courseware use groups K.**

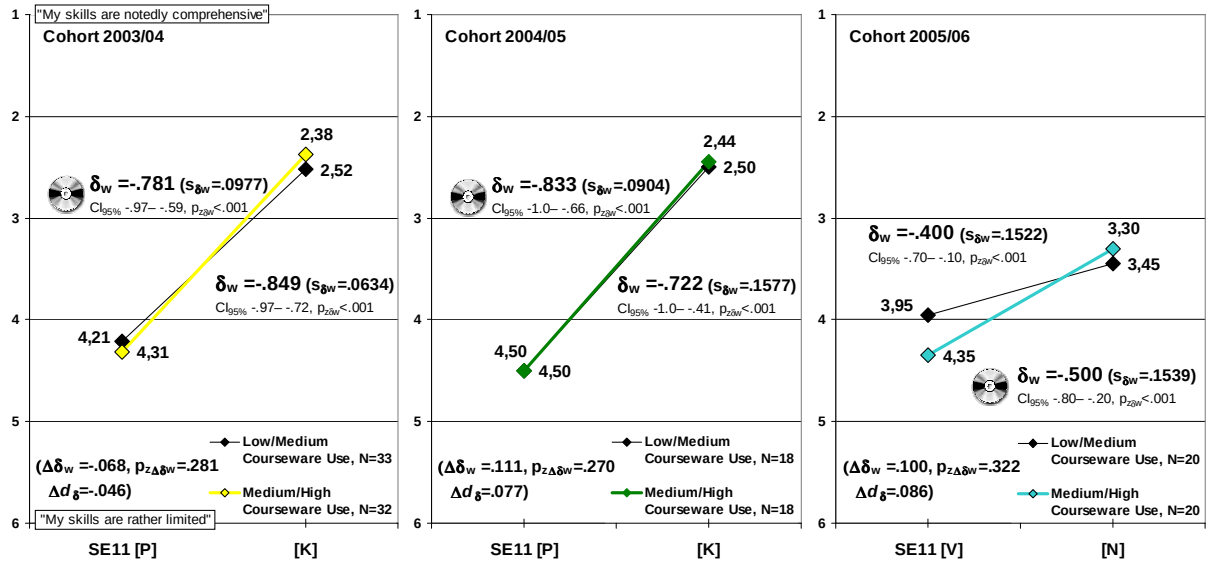


Figure 36. **Changes in conflict moderation process self-efficacy levels [SE11] for median-split ex-post facto courseware use groups K.**

The figures indicate clearly that there is a significant net change of self-efficacy levels in both groups with post-module levels being significantly higher than prior levels. For example, as depicted in Figure 35 for the "High-Courseware Use" group in cohort 2003/04, the 'net' probability for self-rated self-efficacy levels to be superior after module participation is almost 91% (δ_W=.906, p<.001). For the items under scrutiny, this applies to all groups and cohorts.

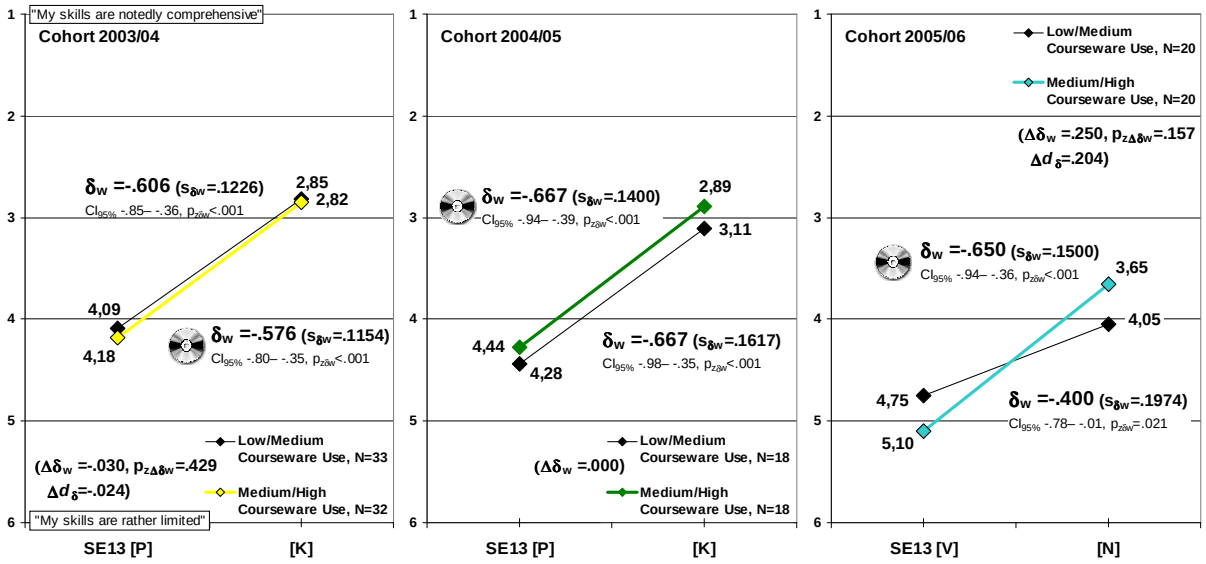


***Figure 37*. Changes in conflict resolution in groups applied skills self-efficacy levels [SE13] for median-split ex-post facto courseware use groups K**

Furthermore, the extent of self-efficacy gains in the "high usage" groups can be compared to that found for "low usage" groups. Differences in increase were either found to approximate zero ($|\Delta\delta_w| \leq 6.8\%$), meaning that "low" and "high use" groups change in similar fashion, or, as in the majority of cases, the increases found for "high use" groups were of greater magnitude ($\Delta\delta_w > 10\%$). For example, as detailed in Figure 37, the differences between net probability of superiority changes found in cohort 2005/06 amount to 25.0% (or, if expressed in terms of Cohen's d_δ .204 effect size units). However, the only difference between group gains significantly ($p<.05$) differing from zero was found in cohort 2005/06 for Mediation Strategy [SE10] ($\Delta\delta_w=.350$, $p=.048$, see Figure 35). This may be taken as an indication that the power of these comparisons may not be high enough to render differences possibly existing between the two groups significant.

Median-split ex-post facto courseware usage groups were also defined for the *Mediation Skills* courseware (i.e. M_w in Table 25 in section 4.2), allowing for two kinds of further analyses. First, in a similar fashion to the foregoing analyses, the net changes in self-efficacy levels can be calculated and contrasted for both M_w groups. And, second, the effects of the combined employment of both coursewares can be evaluated.

As depicted in Figure 38 below, the first class of analyses yield results similar to the *Conflict Resolution in Groups* courseware analyses reported above. There is a significant net change of self-efficacy levels in both groups, with post-*Mediation Skills* Module [M] levels being significantly higher than after the previous module (*Problem Solving*, [P]). And, yet

again, the extent of self-efficacy gains in the "high usage" groups is higher for Mediation Competencies Knowledge (SE07, $\Delta\delta_w$=.167, p=.159) and Mediation Strategy Knowledge (SE12, $\Delta\delta_w$=.167, p=.229). Interestingly, however, in item SE13 – where self-efficacy pertains to Applied Two-Party Mediation Skills rather than practical knowledge – lower increases were found for the "high use" group ($\Delta\delta_w$=-.222, p=.150). The post-module between-group difference is not found significant, however (δ=-.207 with $CI_{95\%}$ between -.521 and +.158 and p=.131), a finding that also applies to the other two items.

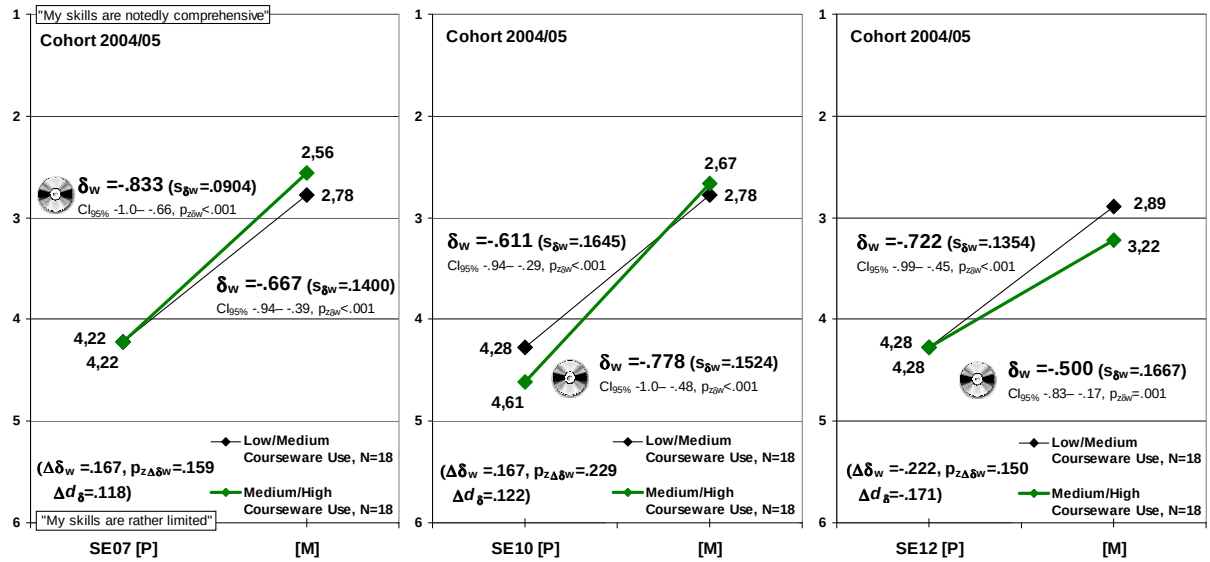


Figure 38. **Changes in self-efficacy levels for mediation competencies (SE07), mediation strategy knowledge (SE12) and applied two-party mediation skills (SE13) in the 2004/05 blended learning *Mediation Skills* Module for median-split ex-post facto *Mediation Skills* courseware use Groups M.**

Based on the defined usage groups for the *Mediation Skills* courseware (M_w) or the *Conflict Resolution in Groups* courseware (K_w), the combined effect of both conflict resolution training coursewares in 2004/05 was also evaluated. In a first approach, a four-group comparison was attempted, based on the four alternative combinations for M_w and K_w, respectively ("Low-Low", "Low-High", "High-Low", "High-High"). In addition, a combined "medium" group was defined for those participants previously subsumed under exactly one "high" usage category in either M_w or K_w. The results for this analysis are reported in Appendix F.1. Here, generally, no interpretable differences were found between the "high/low" or "low/high" or the medium" use groups on the one hand and the "high/high" use group on the other. Thus, a second approach was taken to address the combined effect. Both peer-student tutors as well as some participants had repeatedly put forward their opinion that learning could be enhanced by further reducing the number of similarities between the coursewares. Subscribing to this notion of "excessive similarity" between the two coursewares, it was

decided to define a single "high use" category, subsuming all sample base participants that had used either or both of the coursewares to a high degree ("high" in either K_w or M_w or both). With the number of participants being small in the combined "low use" group, further restrictions in terms of power were expected. To counter this problem, all available intra-participant data across time was included in the analysis. This means that, for each participant on item level, all data available for points of time prior to the *Mediation Skills* module were paired with all data received after conclusion of the *Conflict Resolution in Groups* module. For example, sample base participants of the 2004/05 cohort were asked to respond to item SE 13 in the pre-test [V], the post-test [N], and all post-module online surveys [F, P, M, K]. For item SE13 and for each participant, therefore, six pairs of data can be provided, i.e. [V]—[K], [V]—[N], [F]—[K], [F]—[N], [P]—[K], and [P]—[N]. Treating each pair as a single case leads to a sextupled total number of cases for pre-post comparisons and, thus, to enhanced statistical power. Furthermore, this procedure is likely to contribute to the internal validity of the inferences as it is based on a greater number of points in time which may be expected to substantially reduce the possible impact of chance. The results of these analyses are depicted in Figures 39 to 41 below.

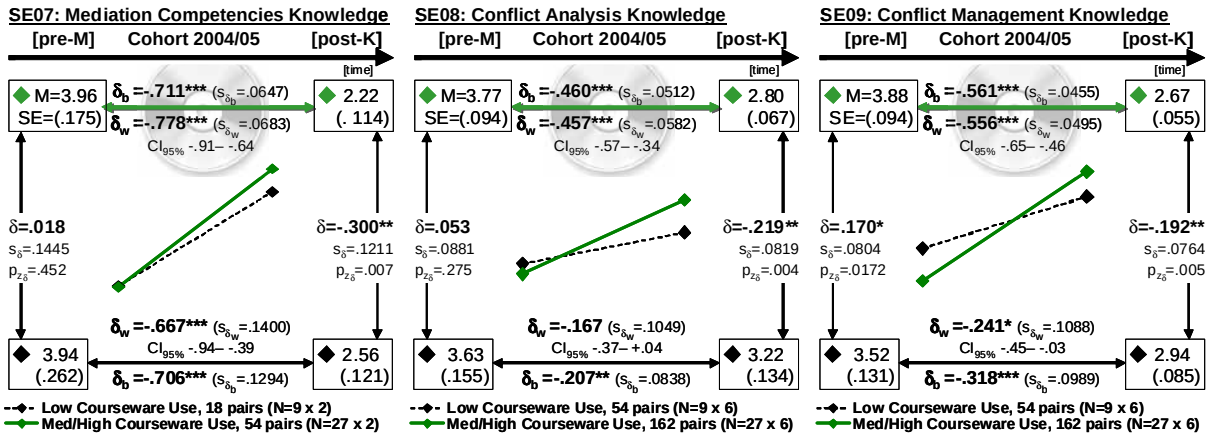


Figure 39. **Changes in self-efficacy levels for mediation-related courseware use group combination (I).**

In Figures 39 and 40, changes in knowledge-based self-efficacy levels are reported. Across items, both the eventual mean self efficacy level as well as the pre-post 'net' gains in self efficacy are higher in the group of those participants having thoroughly used at least one of the mediation-related coursewares. In item SE 10 (Mediation Processes Knowledge), for example, 55.5% of the 18 cases included in the analysis of the "low usage" group report self efficacy gains, whereas 5.5% report a lower self-efficacy at some post-module survey point in

time, resulting in a decrease of scores (i.e. self-efficacy gains) for δ_w=-(55.5-5.5=).500 (or 50%) 'net' of the pairs. This may be compared to a δ_w=.778 in the 'Med/High' group.

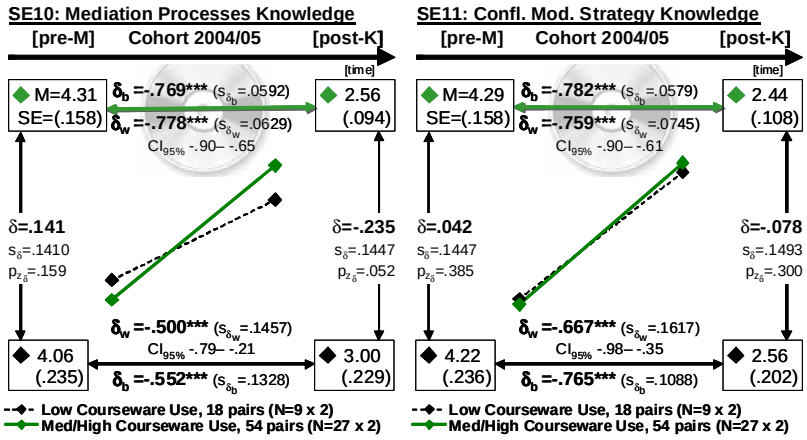


Figure 40. **Changes in Self-Efficacy Levels for Mediation-Related Courseware Use Group Combinations (II).**

Significant ordinal differences for post self efficacy level were found in items 7 (Mediation Competencies Knowledge, δ=-.300, p=.007, Cohen's d_δ=-.328 with CI95% from -.518 to -.063), 8 (Conflict Analysis Knowledge, δ=-.219, p=.004, Cohen's d_δ=-.248 with CI95% from -.396 to -.067), and 9 (Conflict Management Knowledge, δ=-.192, p=.006, Cohen's d_δ=-.220 with CI95% from -.362 to -.073). Significant differences between self-efficacy gains were found in SE-items 8 (Conflict Analysis Knowledge, $\Delta\delta_w$=.290, z_Δ=2.419, p=.008), 9 (Conflict Management Knowledge, $\Delta\delta_w$=.315, z_Δ=2.634, p=.004), and 10 (Mediation Processes Knowledge, $\Delta\delta_w$=.278, z_Δ=1.749, p=.040). For purposes of comparison, the more traditional way of analysis as reported in Table 37 below yields similar results with significant F values in the same items 08 (F(1)=6.873, p=.009, η_p^2=.031), 09 (F(1)=9.677, p=.002, η_p^2=.043), and 10 (F(1)=4.148, p=.045, η_p^2=.056).

Summarizing the above, these findings lend some support to the notion that using the coursewares for preparation may result in augmented self-efficacy change and, eventually, higher knowledge-based self-efficacy levels.

Table 37

Repeated Measures MANOVA for Self-Efficacy Changes in Mediation-Related Modules and Mediation-Related Courseware Ex-Post Facto Groups for Cohort 2004/05

Item		Type III SS	df	MS	F	p	η_p^2
SE07	SE-Level Change (pre-M/post-K)	66.113	1	66.113	67.735	.000	.492
	SE-Level Change x Use Group	.836	1	.836	.856	.358	.012
	Error	68.324	70	.976			
SE08	SE-Level Change (pre-M/post-K)	38.371	1	38.371	41.276	.000	.162
	SE-Level Change x Use Group	6.390	1	6.390	6.873	.009	.031
	Error	198.941	214	.930			
SE09	SE-Level Change (pre-M/post-K)	64.445	1	64.445	76.181	.000	.263
	SE-Level Change x Use Group	8.186	1	8.186	9.677	.002	.043
	Error	181.034	214	.846			
SE10	SE-Level Change (pre-M/post-K)	53.481	1	53.481	66.369	.000	.487
	SE-Level Change x Use Group	3.343	1	3.343	4.148	.045	.056
	Error	56.407	70	.806			
SE11	SE-Level Change (pre-M/post-K)	83.565	1	83.565	88.086	.000	.557
	SE-Level Change x Use Group	.231	1	.231	.244	.623	.003
	Error	66.407	70	.949			
SE12	SE-Level Change (pre-M/post-K)	177.038	1	177.038	224.158	.000	.512
	SE-Level Change x Use Group	1.704	1	1.704	2.158	.143	.010
	Error	169.015	214	.790			
SE13	SE-Level Change (pre-M/post-K)	91.309	1	91.309	10.844	.000	.320
	SE-Level Change x Use Group	.309	1	.309	.341	.560	.002
	Error	193.765	214	.905			

Notes. Sphericity was assumed for within-subjects effects. Two-tailed p-values are reported. Within-subjects linear contrasts are equivalent to the within-subjects effects detailed. Box's Equality of error variance across groups test was not found significant for items SE07 trough SE11 but significant for SE12 (Box's M=10.04, F(3,159790.3)=3.297, p=.020) and SE13 (Box's M=28.93, F(3,159790.3)=9.500, p<.001). Levene tests of error variance homogeneity were not found significant for items SE07, SE08, SE10 through SE13 at both pre-M and post-K as well as for SE09 at pre-M. However, error variance equality cannot be assumed for item SE09 at post-K (F(1,214)=6.420, p=.012) nor for item SE13 at neither pre-M (F(1,214)=17.695, p<.001) nor post-K (F(1,214)=10.004, p=.002).

As for the skill-related self-evaluations, no significant differences could be detected between the ex-post facto groups. There are only small effects for resulting group differences at post-K in item SE 12 (Level of Two-Party Mediation Applied Skills, δ=-.122, p=.063, Cohen's d_δ=-.145 with CI95% from -.304 to +.045). Both graphs in Figure 41 depict changes of the score means. They seem to suggest that the extent of self-efficacy gains are

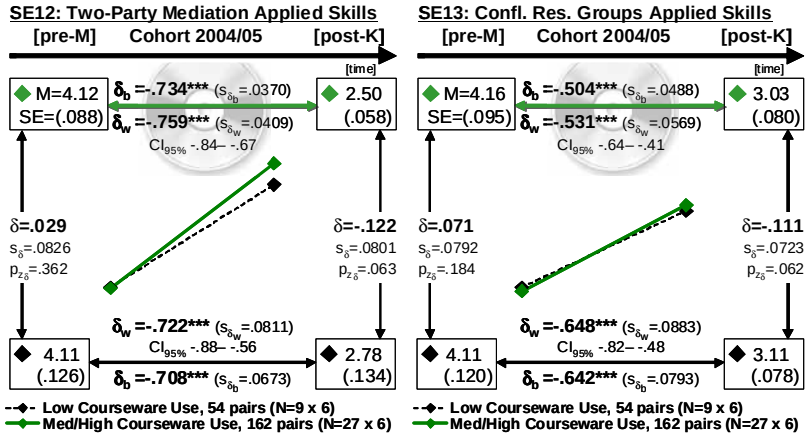


Figure 41. Changes in Self-Efficacy Levels for Mediation-Related Courseware Use Group Combinations (III).

somewhat higher in the "Med/High" courseware use groups. However, detailed ordinal analyses reveal that in item SE13 (Level of Conflict Moderation Applied Skills), post-K self-efficacy levels are higher in only 66.7% of the cases and lower in 13.6% of the cases, resulting in $|\delta_w|$=(66.7-13.6%=).531. This compares to 74.1% and 9.3%, for pre-post-pairs in the "low" use group, respectively, with $|\delta_w|$=(74.1-9.3%=).648. The difference between the gain probabilities is not significant ($\Delta\delta_w$=.117, z_Δ=1.116, p=.132). Neither were significant effects found for courseware use in the repeated measures MANOVA as reported in Table 37.

H7: Preparation and Self-Efficacy Levels.

It was communicated to participants that they should use the coursewares as an elucidatory means of preparation, complementing the more traditional module preparation by reading recommended texts. It was hypothesized that, in doing so, they would profit more from subsequent soft-skills training, resulting in both higher post-module self-efficacy levels (as compared to other participants) and higher self-efficacy gains.

Extensive analyses have been run to shed light onto these assumptions. For items [SE07] to [SE13], self-efficacy levels and pre-post-changes were calculated for the blended learning sample base cohorts (i.e. 2003/04, 2004/05 and 2005/06), divided into four ex-post facto groups per item according to the self-reported use of the *Conflict Resolution in Groups* courseware (K_w, see Table 25 in section 4.2) and the use of suggested reading (textbooks/literature) for both the *Conflict Resolution in Groups* and *Mediation Skills* modules (Lit_{B3}, see Table 28 in section 4.2). Moreover, for each of the four groups and each item, self efficacy levels and level changes were calculated by use of the power-increasing approach described in the previous section. This means that all measurements taken previous to the *Mediation Skills* module (e.g. pre-test [N], online surveys for the *Facilitation* [F] and *Problem Solving* [P] modules, denoted 'pre-M' in the following) were paired with all measurements taken after the conclusion of the *Conflict Resolution in Groups* module (i.e. *Conflict Resolution in Groups* module online survey [K] and post-test, denoted 'post-K' hereafter), treating each pair as a single case.

Contrary to expectation, this analysis did generally produce rather indefinite findings for all module-related self-efficacies. There is one noteworthy exception, however, which relates to the very self-efficacy item essentially targeted at in the *Conflict Resolution in Groups* courseware, namely SE10 (Mediation Strategy/Process Know-How). For this item, the findings

depicted in Figure 42 below are rather consistent across groups and the three blended learning cohorts.

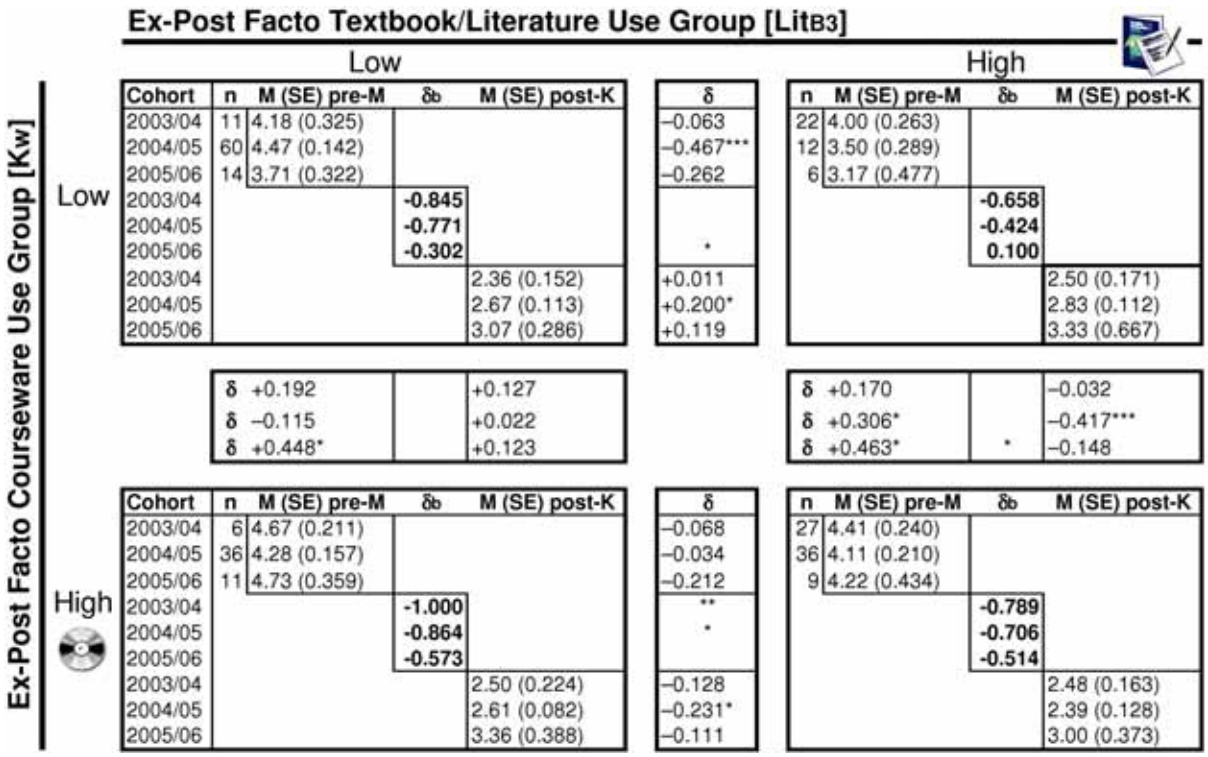


Figure 42. Preparational modes and changes in self-efficacy level of mediation process knowledge (SE10).

The findings may be summarized as follows:

(a) Prior to the commencement of the mediation-related *PROCON* modules, the self-rated level of mediation strategy know-how is significantly lower in those groups subsequently reporting more extensive courseware use.

For example, as depicted in Figure 42, comparing all three cohorts in the "high textbook use" groups, ordinal δ ranges from δ=+.170 to +.463. Across cohorts and textbook use groups, the mean probability for lower self efficacy levels in "high courseware use" groups is 62.2%, and 37.8% for higher levels, resulting in an average δ=(62.2-37.8%=).244 (mean Cohen's d_{δ}=.389 with average $CI_{95\%}$ ranging from -.121 to 1.036).

(b) The groups making solely extensive use of written texts for preparation, i.e. without concomitantly using the courseware, display lower self-efficacy gains than all other comparison groups.

The gains for this group range between δ=-.658 to +.100, with the latter positive value indicating self-efficacy declines. Note that lower scores indicate higher self-efficacy levels as the Likert scale ranges from 1 ("my skills are markedly comprehensive") to 6 ("my skills are rather limited"). Across cohorts, the difference in gains in comparison to all other groups amounts to a mean δ_b=.380 (with an average Cohen's d_{δ}=.596, $CI_{95\%}$ from -.045 to 1.904).

(c) The "high textbook/low courseware" groups also tend to attain lower post-module self-efficacy levels than the other groups. In comparison to all other groups and across

cohorts, the average net probability that self-efficacy levels are lower after completion of the mediation-related modules (post-K) is (a non-significant) δ=.117.

For the vast majority of the other self-efficacy items, no differences in self-efficacy levels were found between groups previous to the commencement of the mediation-related modules (pre-M) that could possibly explain or predict use of the preparational materials offered. While the analyses seem to suggest that some self-efficacy gains may be higher when textbooks were highly used in addition to the courseware, other self-efficacy gains were found significantly higher for those using neither courseware nor textbooks to a noteworthy extent. While the findings obtained in cohort 2005/06 generally lend some support to the hypotheses under question, the findings from previous cohorts seem to contradict these, or vice-versa. In short, specifics may be of interest in subsequent research studies, and the findings are therefore detailed in Appendix F.3. In sum, however, the general picture is far too mixed to allow for making valid statistical inferences concerning the issues under review.

H8: CSSL and 'Accuracy' of Self-Reported Self-Efficacy Levels.

In the foregoing, empirical results as to the possible associations between courseware use and the reported self-efficacy levels were reported. It was hypothesized that courseware use does not necessarily lead to a higher level of self-efficacy. It may be true that awareness is raised in users for the standards and criteria according to which experienced mediators rate the (professional) behavioural models presented in the coursewares. Put in another way, users may also be trained in the application of these standards, and they may become more critical in regard to their own perceived skills. In this case, however, there may be an improved congruence between their subjective self-evaluation and external ("objective") skills evaluations.

In a straightforward statistical approach to this issue, correlation coefficients between reported self-efficacy levels and the knowledge and situational judgment post-tests were calculated for the ex-post facto groups based on courseware use. The results obtained as reported in Appendix F.2 are mixed. Generally, the association was found to be low in most cases, not differing significantly from zero. On average and across cohorts, the strongest associations between self-efficacy level and test scores were found for items SE07 (Mediation Competencies Knowledge) and SE10 (Mediation processes knowledge). And for these items, the correlations generally follow the hypothesized direction, i.e. with higher self-efficacy denoted by lower scores, the coefficients are negative for correlations between self-efficacy

and knowledge scores and situational judgement test "correct" scores, and they are positive for error scores. In many cases, they are also stronger for the "Med/High" courseware use-groups. In sum, however, the associations found are too weak to allow further conclusions.

DIGEST 4.4

Empirical findings were presented investigating the relationship between CSSL courseware use and self-efficacies, i.e. the self-rated confidence in theoretical understanding and mediation-related knowledge as well as in the application of socio-communicative skills utilized in mediation.

No consistent associations were found between self-efficacy gains and courseware use indices. In between-group analyses for knowledge-related self-efficacies, either equal or higher gains and either equal or higher post-module levels were found for the "high" courseware use groups. Across cohorts and items, the effects found are rather small (Cohen's *d* typically between +.0 to +.3), however. As for skill-related self-efficacies, no significant between-group differences in either gains or post-module levels were detected.

Generally, neither corroborating nor contradicting evidence was found with regard to the assumption that combining CSSL courseware and traditional textbook use as means of module preparation positively affects module-related self-efficacies. A consistent picture emerged for the self-efficacy level most closely related to the *Conflict Resolution in Groups* courseware in terms of content. Here, the "high textbook/low courseware use" combination was found to result in lower self-efficacy gains ($d \approx .6$) and lower post-module levels ($d \approx .2$).

Accuracy of self-rated confidence levels was estimated by comparing "low" and "high" courseware use groups with regard to the respective association between the self-rated confidence levels and post-test scores. For the majority of items, the associations found are weak, and differences found vary between cohorts and/or across test scores. Rather consistent between-group differences in line with the stated hypothesis were only found in two self-efficacy level items closely related in terms of content ($\Delta r_s \approx .1$ to .5).

5. Discussion

Theoretical considerations and empirical findings were put forward which gave rise to the idea that Computer-Supported Social Learning (CSSL) technology may enhance training programs for soft skills.

More specifically, eight summative assumptions were detailed above which, in short, presuppose that augmented increases and higher resulting levels of socio-communicative competence may be effected by introducing CSSL courseware to extant soft skills trainings to be used as a supplementary means of preparation.

Among other evaluative questions, the eight assumptions were investigated within the context of a university-based, skills-oriented mediation training programme. Across four consecutive years and participant cohorts, conflict-resolution-related CSSL courseware had been gradually implemented into the curricular modules.

In the previous section, findings were reported concerning the impact: Is the extent of CSSL courseware use associated with self-reported learning gains, knowledge and situational judgement test scores, and perceived confidence in one's knowledge and skills?

In the following, the findings are summarized and evaluated. Factors potentially limiting the validity and generalizability of the conclusions drawn are discussed and potential rival hypotheses are explored. As this study essentially is "proof-of-principle"-oriented, possibly fruitful directions of further CSSL research are also outlined, and possible implications for theory and practice are considered.

5.1 Conclusions

The assumptions evaluated essentially centred around the possible impact of Computer-Supported Social Learning (CSSL): Can augmented increases and higher resulting levels of socio-communicative competence be expected when tailored CSSL courseware is provided as a supplementary means of preparation for soft-skills training?

The impact of courseware use on participants' knowledge, skills, and attitudes was studied in a broader evaluation project undertaken within the context of a university-based, skills-oriented mediation training programme. Across four consecutive years and participant cohorts, conflict-resolution-related CSSL courseware had been gradually implemented into the curricular modules of the programme. The main findings will be summarized and interpreted briefly in the following.

Situational Judgement.

Assessing situational judgment skills was discussed as one way of 'tapping' an individual's socio-communicative competence. Across all cohorts and scoring methods, large positive effects were found for situational judgement test scores. The overall findings also lend support to the assumption that CSSL use effects superior performance levels in post-curricular situational judgement tests (SJT). Across cohorts, significant medium associations were found between courseware use and SJT 'pick-best' scores, and large post-test differences were found between 'low' and 'medium' (or 'partial') use groups. These seem to be strongest where CSSL courseware was used as a complementary, not substitutional means of preparation (i.e. "medium" or "high" textbook use). Augmented score gains were found in a revised, potentially more powerful SJT version in the 2005/06 cohort; the difference in increase between courseware use groups can be regarded as a medium effect.

Attitudes: 'Self-efficacy' and Interest.

Attitudes towards self are measured in many studies investigating the effects of technology-enhanced learning; here, individuals are asked to self-rate their perceived confidence in their own skills and accessible knowledge-for-action. Hoping that these 'self-efficacy' ratings provide valuable clues as to the level of socio-communicative competence, they were repeatedly solicited from participants across time.

"Medium" to "large" effect sizes in 'self-efficacy' change levels were found; these were essentially linked to particular *PROCON* modules. However, across items and cohorts, no

consistent associations were found between self-efficacy raw gain scores and courseware use indices, and across cohorts and items, the effects detected in comparative analyses between courseware use groups may be considered small, at best.

Moreover, neither corroborating nor contradicting evidence was found with regard to the assumptions that courseware use may positively effect the 'accuracy' of self-rated confidence levels or that module-related self-efficacies are positively affected by the combination of CSSL courseware and traditional textbook use as means of preparation.

Similarly, another form of attitudes, namely, the post-curricular level of self-rated interest in the subject matter (i.e. mediation, conflict resolution) was not consistently found higher for "high" CSSL courseware use groups.

Conceptual Knowledge.

Conceptual knowledge was characterized as a prerequisite for socio-communicative competence which traditionally is imparted by use of written materials and by means of lecturettes in training sessions.

Between pre- and post-tests, significant gains in conceptual knowledge test scores were detected; the magnitude of effect may be characterized as "medium". More specifically, it was hoped that additional courseware may positively affect knowledge of the concepts and the taxonomy of the mediation model presented in the curriculum.

The findings detailed above generally lend support to this notion, but the impact was found weak, too indefinite and possibly restricted to 'thorough' users of the coursewares.

DIGEST 5.1

The impact of the introduction and use of CSSL courseware on participants' socio-communicative competence was studied in a broader evaluation project undertaken within the context of a university-based, skills-oriented mediation training programme.

The findings lend support to the assumption that preparatory CSSL use positively influences gains and resulting ability to judge inter-personal situations and behaviours. No significant effects were found for self efficacy gains and levels or for conceptual knowledge.

5.2 The Validity of Statistical Conclusions

Four independent consecutive participant cohorts were included in this study, and where possible, statistical inferences were made on the basis of the greatest number of available data. To forestall statistical conclusion invalidity due to 'fishing' phenomena, inferences made were principally limited to such findings where the direction of associations or between-group differences was fundamentally consistent across all subgroups included in an analysis. Deviations from this principle (e.g. contradictory effects found) were explicitly detailed. Moreover, inferences made were preferably based on composite indices and various scoring algorithms rather individual items and calculations, and on repeated rather than single comparisons. An ordinal level of measurement was assumed for most items and instruments and where salient violations to non-normality could not be ruled out, robust and distribution-free statistical methods were used for analysis. For these reasons, inferences based on significant findings may be viewed as reasonably justified. The effect sizes reported are also largely based on ordinal comparisons, re-interpretations of measures of association, and dominance statistics. In principle, therefore, estimates of magnitude should be treated and interpreted with due prudence. Possibly, to be on the safe side, effect sizes should also be compared in ordinal terms.

Of the various effect sizes detected, the highest magnitudes of effects were found for the situational judgement variables. Empirically speaking, therefore, these may lend the strongest relative support for the notion that the introduction and learner's use of tailored CSSL courseware positively effects learning gains and resulting levels of socio-communicative competence in areas addressed by the 'blended delivery package'. Associations in the hypothesized direction were found between courseware use and SJT post-test scores were found across scores and across cohorts. As, statistically speaking, group-based comparisons may be generally even more powerful, it does not seem surprising that even partial or medium courseware use yields significant effects.

Significant differences in group means and/or distributional overlaps were also found for learning gains and changes of self-efficacy level raw scores. By contrast, however, effect sizes for associations between courseware use and self-reported learning gains as well as changes of self-efficacy level raw scores generally were relatively low.

This may be due to low statistical power. The results of analyses based on associations require reliable instruments, whereas reliability in the dependent variable (not in the difference or gain itself) may likely be of subordinate importance in analyses of differences in experimental approaches (Nicewander & Price, 1983; Humphreys & Drasgow, 1989; May & Hittner, 2003). Results of a test-retest estimate for self-efficacy levels repeatedly reported by participants across time are detailed in Table 38 below.

Table 38

Cross-cohort Associations between Self-Reported Self-Efficacy Levels prior to PROCON *Modules with Corresponding Subject Matter (Item Reliability Estimation)*

SE-Item	Associations included [instruments / number of respondents]	Z-averaged r_s (unweighted)	Z-averaged r_s (weighted)
07 Mediation Competencies Knowledge	F-P (42)	.467	.467
08 Conflict Analysis Knowledge	V-F (43), V-P(207) , F-P (42)	.431	.433
09 Conflict Management Knowledge	V-F (43), V-P(207) , F-P (42)	.464	.483
10 Mediation Strategy Knowledge	F-P (42)	.488	.488
11 Conflict Moderation Knowledge	F-P (42)	.373	.373
12 Mediation Skills	V-F (43), V-P(207) , F-P (42)	.456	.534
13 Conflict Resolution in Groups Skills	V-F (43), V-P(206) , F-P (42)	.495	.539
07-13 (all items / mean)	all of the above	.476[a]	.395[b]
	(95% CI based on Z_r)	.288 – .628	.347 – .441

Notes. Confidence interval with limits $(e^{2Z_r}-1)/(e^{2Z_r}+1)$ based on Zar's (1999) correction for r: $Z_r=\ln((1+r)/(1-r))/2 \pm z_\alpha\,(1.060/(k-3))^{1/2}$. [a]Based on all association coefficients with 95% CI based on average number of participants (k=1293/15). [b]Based on all association coefficients, weighted by the total number of participants across associations (k=1293).

Compared to other procedures used to estimate reliability, correlating test and re-test-scores is a conservative approach. For no variation to occur across time, participants would have to interpret questions always in exactly the same manner and to tick exactly the same resulting scale box. Given the applied context and the exploratory evaluative strategy, however, item wording primarily targeted at obtaining intuitive statements and allowed for variation across time. Item formulation may be considered relatively crude, favouring broadness, comprehensiveness, and individual cognitive integration of perceived item indicators (*Indikatorverschmelzung*) over cross-participant consistency of question comprehension (cf. Langer & Schulz von Thun, 1974). Given this context, it is not unreasonable to argue that there may be a relatively low probability that, on 6 to 15-point ordinal and discrete Likert scales, an individual participant retains exact score levels across time.

Test-retest reliability estimation is also based on the assumption of stable true scores. By contrast, one may not rule out that learning in related areas such as problem solving or facilitation methods also affect mediation-related self-efficacy reports obtained prior to the

commencement of the mediation- or conflict-resolution-related learning modules. The 15 associations calculated across self-efficacy items and cohorts were found to range between r_s=.279 and r_s=.622. Across items, the weighted mean association is found to amount to ≈.4. This association – which may also be termed 'stability' – may be viewed as relatively low when compared with, for example, available intelligence or personality trait tests. The degree of absence of stability may also be expressed in terms of residual variance or "variability non-overlap", as $1-r^2$, which in the case of the weighted average association across items is ($1-.395^2$=).844, or 84.4% .

By contrast, 'stability' for the distributions as a whole (the parameters of which are used to calculate median and arithmetic mean) included in Table 38 is relatively high. In Table 39 below, ordinal between-group 'net' differences are reported in terms of Cliff's δ_b. This is also an estimate of non-overlap (Romano et al., 2006), which is the issue here, and therefore, unweighted and weighted averages for δ_b (across comparisons and items) were calculated on the basis of absolute values for δ_b.

Table 39

Cross-cohort Ordinal Dominance Self-Reported Self-Efficacy Levels prior to* PROCON *Modules with Corresponding Subject Matter (Distributional Stability Estimation)

SE-Item	Included Comparisons [instruments / number of respondents]	Averaged δ_b (unweighted)	Averaged δ_b (weighted)
07 Mediation Competencies Knowledge	F-P (42)	-.228	-.228
08 Conflict Analysis Knowledge	V-F (43), V-P(207), F-P (42)	-.117	-.004
09 Conflict Management Knowledge	V-F (43), V-P(207) , F-P (42)	-.094	.008
10 Mediation Strategy Knowledge	F-P (42)	-.199	-.199
11 Conflict Moderation Knowledge	F-P (42)	-.264	-.264
12 Mediation Skills	V-F (43), V-P(207) , F-P (42)	.369	.129
13 Conflict Resolution in Groups Skills	V-F (43), V-P(206) , F-P (42)	.029	.034
07-13 (all items / mean)	all of the above	.129[a]	.094[b]

Notes. [a]Based on the absolute values of all 15 ordinal dominance coefficients. [b]Based on the absolute values of all 15 ordinal dominance coefficients, weighted by the number of respondents.

The maximum absolute value vor the δ_b measure is .264, which corresponds to saying that the highest (ordinal) non-overlap between the distributions found is .138 (or 13.8% when expressed as percentage). The δ_b averages also lend some support to the notion that distributional stability is high (δ_b=.129 signifies a non-overlap of 6.89%, and for δ_b=.094, distributional non-overlap is .05, or approximately 5%).

The fact that no consistent significant effects were found for self efficacy gains and levels or for conceptual knowledge may thus be partly due to problems of statistical power

induced by a lack of intra-personal reliability in the self-reports. Future research may therefore, by use of more accurate instrumentation, establish links between CSSL use and self-efficacy.

DIGEST 5.2

To forestall problems of the validity of statistical conclusions, violation of the assumptions of statistical tests were sought to be avoided, and adequate yet powerful statistical methods were used throughout. Moreover, statistical inferences were principally limited to consistent findings, based on composite indices, various scoring algorithms, the greatest number of available data. Thus, inferences based on significant findings may be viewed as reasonably justified. By contrast, while both the stability of the distributions as wholes and of ordinal group differences may be assumed, the lack of intra-personal reliability in self-reported self-efficacy presents a problem thought to considerably reduce statistical power.

5.3 The Validity of Instruments and Inferences

Summarizing the above, there is strong evidence, based on various sources, that students passing through the *PROCON* curriculum derive significant learning benefits. Furthermore, augmented increases and higher resulting levels of situational judgement test scores were found for those groups using Computer-Supported Social Learning (CSSL) courseware designed to provide situational anchors and cues for conceptual knowledge.

In the following, three main questions will be discussed. First, theoretical validity is addressed: Are situational judgement test scores a valid indicator of socio-communicative competence? Or, more pointedly: Were gains found for situational judgement *skills* - or simply for test *scores*? Second, may the gains and superior post-test scores be validly attributed to the introduction and use of complementary CSSL courseware? And, finally, to what extent are the findings obtained in this idiosyncratic environment generalizable to other populations, settings, and across time and space?

Instrumentation.

One must assume an overwhelming complexity of the constructs attempted to be measured. Naturally, this applies to the overall conceptions of "social skills" and the more area specific "socio-communicative competence", but even when constructs are defined in much narrower terms such as, for example, "judgement skills of behaviours and situations of group-based mediation and facilitation", instruments may possibly only probe for certain aspects.

Critics may therefore argue that the instruments and scoring procedures detailed above may not be sufficient to interpret beyond item level. Indeed, it may essentially be a matter of future research in this area to substantiate claims to validity and to provide estimations of trait validity; but it is also true that this may be said of any new test. For the time being, one may define constructs in terms of the available measures, or, as Boring had put it for the intelligence testing domain, "intelligence is what the tests test" (p. 35). Arguing along these lines, socio-communicative competence or mediation-related judgement skills would possibly, but not necessarily show in the situational judgement tests.

By no means, this should be taken as an endorsement of sloppy test construction, however. By contrast, a conscientious creation of measures construction would nevertheless attempt to map and capture relevant areas of the construct's nomological network. This however, is a more general problem of the content validities of both CSSL and situational

judgement research. On the one hand, tailored coursewares may always pertain to a limited number of conceptual models thought to be helpful for coping with a specific set of typical inter-personal situations and defined roles found in a given area or domain of inter-personal communication. Testing simply for conceptual knowledge, not skills, is unsatisfactory. Skills-based testing, however, always refers to situated conceptual knowledge and thus refers to a limited number of situations. With the number of potential target situations and conditions be infinitely greater, the representativeness of situations, behaviours or incidents used in a given test may thus be always be subjected to criticism. On the other hand, selecting a sample of typical situations and critical incidents for test purposes may prove equally unsatisfactory as helpful, tailored CSSL software may acquaint learners with just these. Then, CSSL courseware and situational judgement are simply two sides of the same coin.

As in many evaluation projects, the measures used in this study had to be created from scratch as no instruments were available at the time. The situational judgement tests were aimed at capturing judgment of mediation- and facilitation-related situations and behaviours in not just one but across various stages of the mediation process. While many characteristics of the situations and incidents used in the situational judgement tests did naturally resemble those used in the CSSL courseware, considerable variations were introduced for the background narrative and the behavioural options offered as multiple choices. Moreover, knowledge tests, self-reports, and situational judgement tests were initially developed based on series of interviews with programme management, teaching staff, and peer tutors, conducted after reviewing teaching materials and attending workshops and tutorials. For these reasons, certain degrees of face validities may be inferred.

Generally in line with the paradigms of action research, the instruments were also continually revised and updated to incorporate feed-back from stakeholders. Changes were implemented when they were thought to positively contribute to the psychometric properties of the instruments, especially in terms of measurement validity, and/or when there was no viable alternative to do so. In the majority of cases, single items were added or omitted, or wording was adjusted to conform with the terminology used in a specific cohort. (For example, the specific names of the modules had been changed over time.) Thus, changes may be considered minor in the majority of cases. In two cases, however, major changes were implemented. First, the video-based situational judgement test (SJT) used in the 2002/03 cohort had largely been based on spin-off material from the production of the *Conflict Resolution in Groups*

courseware. In order to avoid possible interference due to priming or halo effects, and, thus, potential validity problems, a parallel SJT version had to be used in subsequent cohorts. And, secondly, with the major aim to improve power, findings from both the evaluation of responses to the SJT used in this study as well as general SJT research were incorporated in a major SJT revision used in the 2005/06 cohort.

On the one hand, these changes likely affect inter-cohort comparisons. On the other hand, most of the results presented above allude to cross-cohort comparisons based on ex facto-groups and parallel tests, or to single-cohort comparisons.

However, due to the lack of exact tests of the psychometric properties of the tests used in this evaluation study, doubts must remain as to the extent to which the nomological networks of the constructs have indeed been captured, and to the generalizability of knowledge and judgement skills to contexts other than the measurement context itself. A related issue of measurement validity pertains to the analogies between the *Conflict Resolution in Groups* courseware and the situational judgment tests (SJT) employed at pre- and post-tests. They are largely based on the same task principles, namely, (a) being confronted with a specific question or task, (b) observing and judging the situational characteristics of video-scenes presented, and (c) deriving an adequate reaction (picking the 'best' option of various reaction alternatives presented). Moreover, in order to fully understand the task, setting and subsequent scenes partially requires knowledge of previous scenes, as the scenes are all based on a background story sequenced as a narrative. In both courseware and SJT, this narrative follows the same underlying stage model of group-based conflict resolution. Thus, critics may argue that what should have been a test of situational judgement skills may actually be more a test of specific courseware knowledge (or ability to recall).

One may counter-argue that – without significantly jeopardizing psychometric parallelity –various differences between the two background stories had been introduced to increase the probability that these very effects are alleviated – including, but not limited to, a complete change of the setting, modifications in terms of the conflicting issues, the sexes of the mediator and the parties in conflict, and the behavioural stance of the team leader and his partners-in-conflict. Furthermore, many SJT item stems and most of the alternatives presented available differ significantly from the courseware tasks – which possibly had led to considerable error variance if courseware knowledge (or the ability of recall) had been the target construct. The fact that a narrative had been used may, despite possible halo problems,

be hypothesized to additionally account for a core skill, namely, an understanding of the mediator's strategy in group-based conflict resolution.

In sum, however, there is currently no empirical evidence that may lend further support to the validity of either these claims. Future research may, for example, provide further SJT tests of group-based mediation scenarios that could be used as external criterion to provide an estimate as to the convergent validity of the SJT. Moreover, in terms of the socio-communicative competence model presented in section 2, the SJT approach to assessing mediation skills should lead to similar results as other approaches (especially performance assessment). It may be a very promising line of future research to test not just the concurrent and predictive validities of situational judgement tests in general (and this SJT in particular), but also to evaluate the strengths, weaknesses, and concurrent validities of all approaches to social skills assessment detailed above.

One final issue of instrumentation refers to the fact that post-test-only-designs were gradually replaced by pre-post-designs. This modification may potentially influence single cohort comparisons as well as between-and cross- cohort comparisons; thus, the potential ramifications of repeated testing need to be addressed. It may be hypothesized that, due to retention and other effects, participants already 'knowing' the situational judgement post-test tend to attain higher scores - independent from any 'real' changes in their abilities to judge mediation-related situations or behaviours. This effectively means that, at least in part, score gains may not be attributable to learning (e.g. by use of courseware), but simply to repeated testing effects.

In the 2004/05 cohort, participants were, at pre-test, randomly assigned to one of two groups either taking one of the two parallel SJT versions. Prior knowledge of post-test items should, if repeated testing effects are present, lead to augmented increases in score gains and, possibly, higher post-test score levels. To test this hypothesis, an extended analysis based on the same data used in Figure 29 (see section 4.3) was performed.

Ordinal differences (in terms of Cliff's δ statistic) were calculated for SJT test scores [MM_SRE] between the two groups tested with different pre-test versions (labelled 'Mozart' and 'Schütz', see. Appendix C., Section III) and the same post-test version ('Schütz'). Despite the somewhat higher pre-test score average in the 'Mozart' group, the two versions were not found to differ significantly (δ=-.059, p=.764, two-tailed; $\Delta M_{[V\_MM\_SRE]}$=-16.8, t(34)=–.60, p=.725, two-tailed). Neither score gains ($\delta_{b(Mozart)}$=.787***, $\Delta M_{(Mozart)}$=+127.6, $\delta_{b(Schütz)}$=.789***,

$\Delta M_{(Schütz)}$=+143.) were found to significantly differ between groups either ($\Delta\delta_b$=.003, p=.989; F(1,34)=.251, p=.620) nor were post-test scores (δ=.053, p=.790, two-tailed; $\Delta M_{[N\_MM\_SRE]}$=-.9, t(34)=–.04, p=.965, two-tailed). In addition, possible confounding influence on courseware use was examined; the results of this analysis are depicted in Figure 43.

Again, there is a main effect for courseware use as all pre-post comparisons were found significant, with δ=1.0 within-participant probabilities for the high courseware groups to attain superior post-test scores. However, no significant differences were detected between pre-test version groups, neither at pre- nor at post-test. The only (weak) indication that repeated testing effects may be present after all comes from a direct comparison of post-test score levels for the two "low courseware use" groups with different pre-test versions. Comparing all post-test scores for these groups individually, 55% of the scores for the "same-post-test" were found to be superior as contrasted to 45% for the reverse, resulting in a (non-significant) net δ=(55%-45%=).100. Again, the post-test levels of "high courseware use" groups are approximately equal.

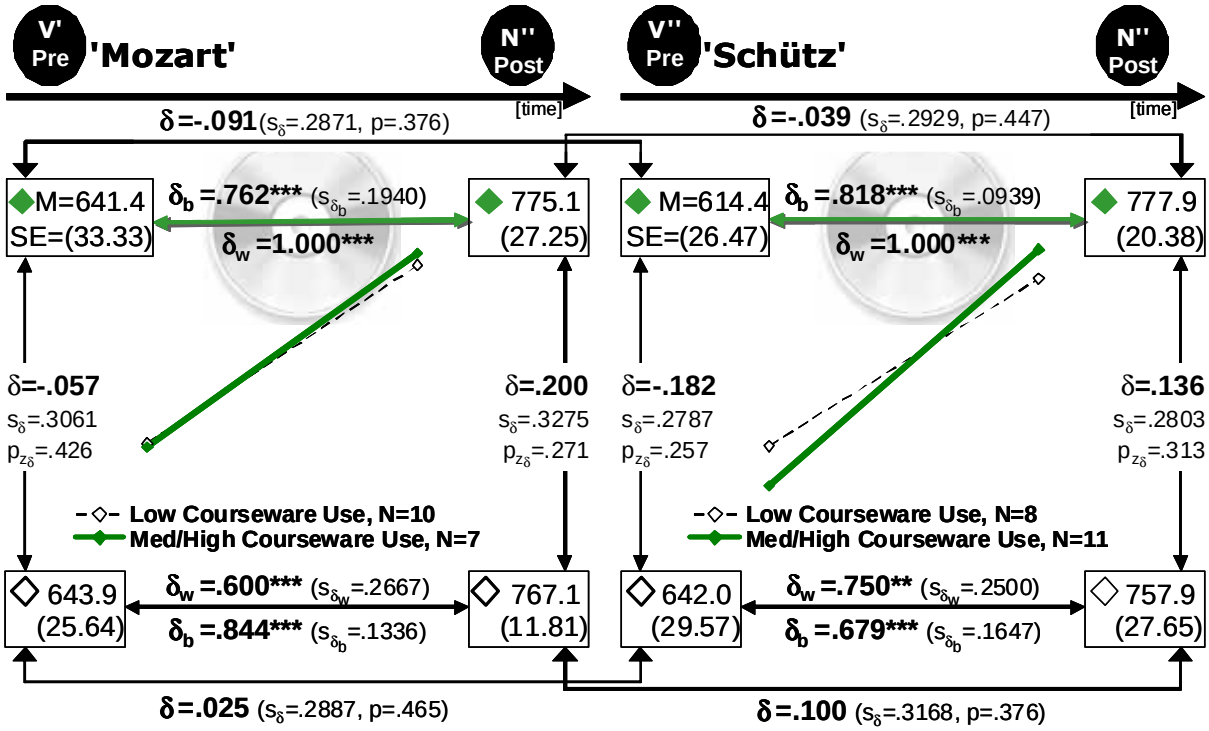


Figure 43. **Courseware use, learning gains, and situational judgement post-test scores [MM_SRE] for two groups with different situational judgement pre-test versions in the 2004/05 cohort.**

In sum, no significant evidence was detected that would lend support to the notion that repeating the situational judgment pre-test at post-test automatically results in higher post-test scores or augmented score gains. One interesting issue remains, however, namely, the question

if pre-testing did have influence beyond the pre-test, namely, on module preparation and classroom-based training. This, however, is essentially a question of the generalizability or external validity of the conclusions drawn.

Design Considerations.

As outlined above, instrument change is one threat to the internal validity of the conclusions drawn. However, further rival hypotheses will be explored in the following that arise from the fact that all evaluative research was based on quasi-experimental settings, rather than a controlled experimental setting which was not feasible (see section 3.4). Inferences, therefore, are largely based on pre-post- or post-test-only comparisons, by use of consecutive student cohorts and ex-post facto groups as proximate quasi controls.

As a consequence of the absence of random assignment, pre-test equivalence neither of the cohorts nor of the ex-post facto courseware use groups may be assumed. In other words, there is increased possibility for differential between-group allocation of confounding factors and error variance. Among the confounding factors to be reviewed, in the following therefore, are those variables found to differ between groups at pre-test.

Moreover, the analyses presented above are largely based on those participants having completed the curriculum within one year (referred to as 'sample base' throughout). It may, however, pose a potential threat to the validity of inferences based on differences between ex-post facto groups if attrition rates can be shown to differ between these groups. This issue will also be further explored in the following.

Selection and Attrition Biases.

The inferences made are largely based on those participants having completed the curriculum within one year (referred to as 'sample base' throughout). In most cases, therefore, students who discontinued participation (either temporarily or completely) were not included. Essetially, this is an ignorability issue; pre-test differences between excluded participants and the sample base in terms of sex, age, prior duration of study, occupational experience and other demographic characteristics were analysed and statistically tested for significance (see Table 14 in section 3.5 and Appendix D.1).

As detailed in Table 15 (see section 3.5) and Appendix D.1, systematic between-group differences were found for three variables, namely

(1) the degree of occupational experience (as measured in years, [V_15, see Appendix A]), with mean decreases across time (i.e. cohorts);

(2) the duration of study prior to participation (as measured in semesters, [V_02, see Appendix A]), with a tendency for senior students to report lower courseware use; and, interestingly,

(3) the sensitivity (*Sensibilität*) score on the standardized *BIP* subtest (Hossiep & Paschen, 1998; [Sen_S], see Appendix A), with a trend for more "sensitive" participants in terms of the *BIP* to report higher courseware use.

All three variables may be hypothesized to be associated with situational judgement skills and with susceptibility for social learning processes (and thus, possibly with courseware-induced learning gains). However, it is not unreasonable to believe that they may also be associated with courseware use. Theoretically, therefore, these variables may confound the inferences presented above.

Table 40

Associations between Potential Confounders and Situational Judgment Test Scores

Posttest Score	Cohorts[a]	Age [V_02]	Occupational Experience [V_15][b]	BIP Sensitivity Score [Sen_S]
N_MM_S1R	2002-04	-.085	-.085	.109
	2005	.029	.107	—
N_MM_S2R	2002-04	-.008	-.067	.072
N_MM_SRE	2002-04	-.026	.041	-.061
	2005	-.261	-.117	—
N_MM_S1F	2002-04	-.072	-.097	.000
	2005	-.234	-.115	—
N_MM_S2F	2002-04	-.154 *	.001	.112
N_MM_S3F	2002-04	.049	.101	-.141 *
	2005	.109	.102	—

Notes. Spearman rank correlation coefficients (*r*). *p<.05, **p<.01, ***p<.001 (one-tailed). [a]Year of pre-test, i.e. 2002-04 based on sample base cohorts 2002/03, 2003/04, and 2004/05 (n=153 and n=130 for V_15), 2005 based on 2005/06 cohort sample base only (n=40).

In a first approach to this potential problem, the relations between the aforementioned variables, courseware use, and the outcome parameters were explored. In Table 40, correlations between the potential confounders and situational judgement post-test scores are detailed. The associations found are weak, ranging between zero and $|r_s| \approx .26$. As groupwise comparisons may potentially offer higher degrees of statistical power, relationships between the three potential confounders, courseware use, and situational judgement scores were explored by use of dichotomized median-split groups. The results are detailed in Figures 44 and 45 below.

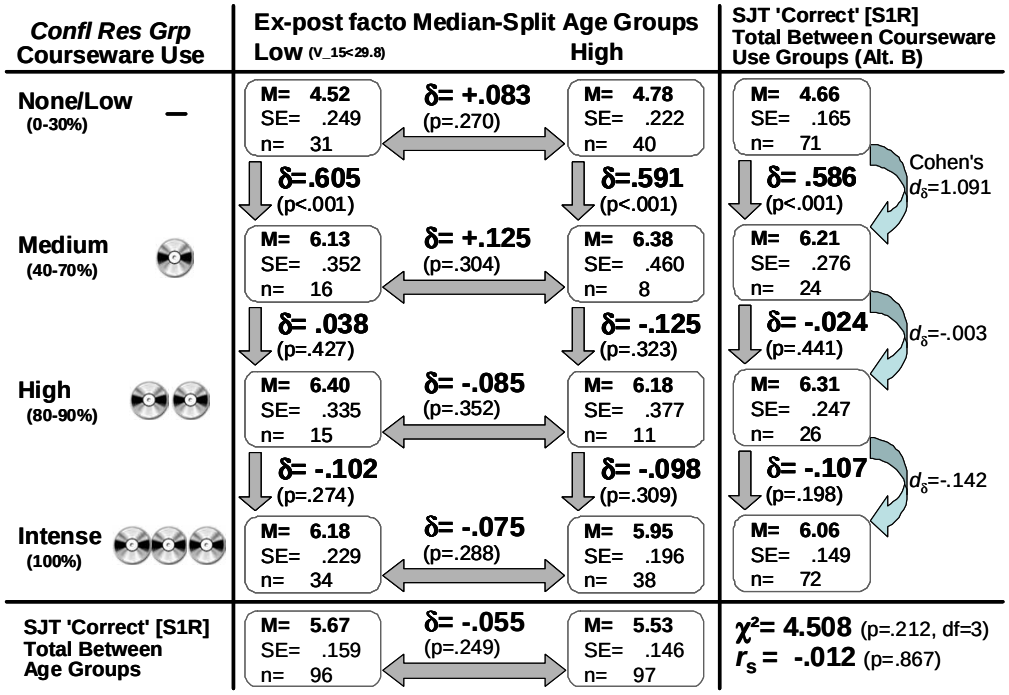


Figure 44. Relationship between age, courseware use, and situational judgement post-test scores.

For purposes of comparison, Figures 44 and 45 draw on the same samples used to create Figure 27 (see section 4.3), and they are structured accordingly. Across all variables and median-split groups, significant differences for situational judgment scores were found between the "no/low" and "medium" courseware use groups (i.e. subgroups 1 and 2 in grouping alternative B as detailed in Tables 26/27, see section 4.2). By contrast, however, neither systematic nor significant relationships or effects of the potential confounding variables were found.

In sum, while these analyses do not rule out the possibility that relations between the variables may exist after all, confounding influences could not be recognized in this study. For the time being, therefore, it seems reasonable to relay this issue to future enquiries into this matter and to explore alternative confounders.

The same holds true for potentially demoralizing effects the courseware or other factors that may – implicitly and uncontrolled for – have influenced attrition rates. It may be true that the attrition phenomena have not been captured by the exploration of pre-test differences between the sample base and excluded cases, as differential attrition may be a biasing factor that potentially poses a threat to the validity of the conclusions drawn.

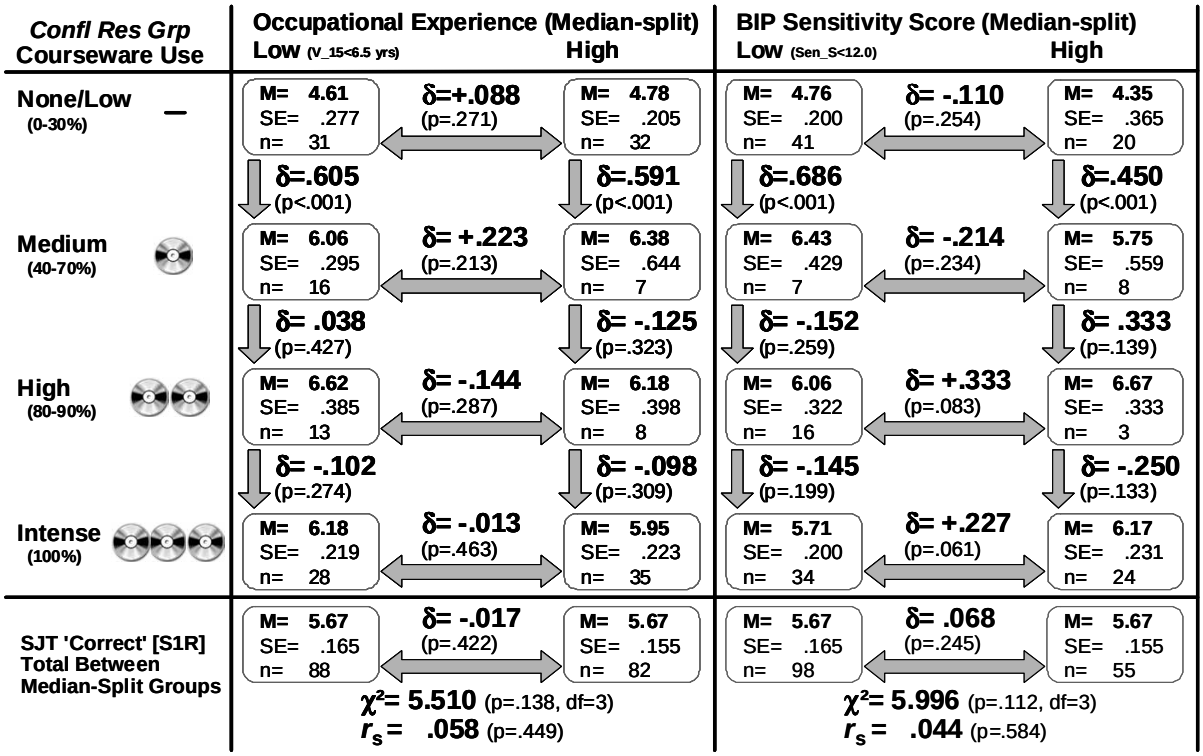


Figure 45. Relationship between occupational experience, *BIP* sensitivity score, courseware use, and situational judgement post-test scores.

In many cases, participants were invited to openly state their grounds for discontinuation. Of the feed-backs received, the vast majority gave personal reasons, seemingly unrelated to any features of either the study or the coursewares. However, in order to effectively identify or rule out interaction effects between courseware use and attrition, motives for dropping-out should have been systematically explored, and, if possible, discontinuing students should have subsequently been asked to report self-efficacy levels or even to partake in post-test examinations to study the differential effects of attrition. This issue, therefore, may also be addressed by future investigators.

Blended Learning Package.

Throughout this study, module preparation and classroom-based training have been viewed as integrated components and have been evaluated as such 'package'. 'Blended learning' researchers as well as practitioners have repeatedly pointed out that computer-based and classroom-based learning mutually depend on each other (e.g. Draper et al., 1997; Moshinskie, 2002; Hess, 2006). For example, the motivation to utilize computer-supported instructional content for learning may also depend on traditionally delivered face-to-face training. Or, put more pointedly, the technology-based part of "blended learning" may also be understood to be classroom-supported. Indeed, despite the fact that the respective courseware

was handed out to participants well before the start of the *Conflict Resolution in Groups* module, many students reported to having used the courseware chiefly between the peer-tutored kick-off session and the subsequent training sessions.

Thus, the "traditionally" delivered face-to-face training and simulation sessions may also be 'confounding' factors in that they potentially exert influence on both courseware use and learning gains. Empirical results indicate that the rated quality of classroom training may indeed be associated with courseware use and both post-curricular situational judgement. The participant's overall rating score for the Conflict Resolution in Groups peer-tutored training course [K_14, see Appendix B.4] was used as an empirical indicator of training quality in Table 41. Significant rank correlation coefficients in the hypothesized direction were also found between most situational judgment test raw score gains and the overall training (e.g., for the 2004/05 cohort with n=36, $r_{s[K\_14]x\Delta[MM\_S2R]}$=.367, p=.014; for 2005/06 with n=40, $r_{s[K\_14]x\Delta[MM\_S3F]}$=-.367, p=.006).

Table 41

Associations between Training Quality, Courseware Use, and Situational Judgment

Variable	**CUI_{TTL}**		**N_MM_SRE**		**N_MM_S1R**		**N_MM_S1F[b]**		**_S2R**	**_S2F[b]**	**N_MM_S3F[b]**	
Cohorts[a]	**'02-'04**	**'05**	**'02-'04**	**'05**	**'02-'04**	**'05**	**'02-'04**	**'05**	**'02-'04**	**'02-'04**	**'02-'04**	**'05**
Training [K14]	***.133****	***.312****	***.180****	***.215***	***.217****	***.134***	***-.241****	***-.229***	***.230****	***-.165****	***-.148****	***-.271****
CUI_{TTL}			.189*	.397**	.449***	.422**	-.017	-.186	.234**	.007	-.179*	-.273*

Notes. Spearman rank correlation coefficients (*r*). *p<.05, **p<.01, ***p<.001 (one-tailed). [a]Year of pre-test, i.e. '02-'04 based on sample base cohorts 2002/03, 2003/04, and 2004/05), '05 based on 2005/06 cohort sample base only (n=40). [b]'Salient Error' scoring, based on the number of occurrences where participants picked (rated as 'best') the alternative least preferred by subject matter experts.

The relationships between the quality of training, courseware use and situational judgement post-test scores were explored further by means of stratification. Two different conditions were examined by use of median-split ex-post facto comparison groups established for the overall classroom-based training quality ratings. The results depicted in Figure 46 may be summarized and interpreted as follows:

- Similar to the findings presented previously, significant differences in post-test scores were found between the "no/low" and "medium" courseware use groups.
- In addition however, a confounding influence of the *Conflict Resolution in Groups* training can be made out as, across usage groups, significant differences were also found between the two training quality groups.

- The impact of training quality may be expected especially when the CSSL courseware had been used to a high or very high degree.

These results lend further support to the idea that there is a mutual interdependence between the virtual and the real, the e-learning and the classroom-based learning. For this reason, evaluative strategies that take account of this interdependence (cf. e.g. Baumgartner & Payr, 1996; Draper, 1997; Draper et al., 1996) may be of more value than studies implicitly viewing e-learning and face-to-face learning as separable entities of "blended learning". While further analyses go beyond the scope of this "proof-of-principle" study and are therefore not reported here, a variety of characteristics of both learners and the social learning environment was simultaneously evaluated in the course of this evaluation project. In future re-analyses, therefore, it may a worthwhile undertaking to further explore this relationship.

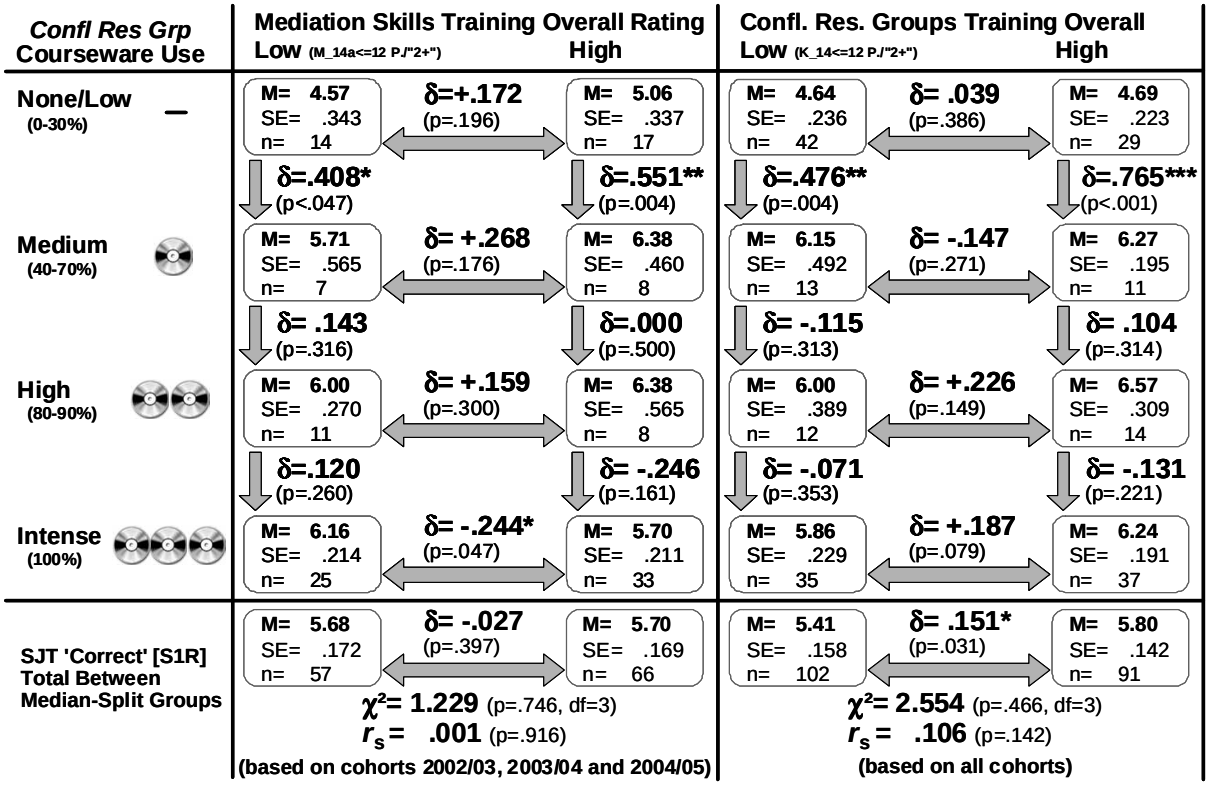


Figure 46. Rated quality of classroom-based training, courseware use and situational judgement post-test scores.

Hawthorne Effects and Generalizability.

This study is one of those in which the effects of introducing innovative computer-supported technologies to extant learning environments were evaluated. For these studies, reviews and textbooks of technology-supported learning and instruction suggest the presence

of further sources of confounding, collectively summarized and labelled as *Hawthorne* effects (cf. e.g. Schulmeister, 2002a).

As R. Olson, Verley, Santos, and Salas (1994) have exemplified, a variety of meanings is associated with this label originally referring to the research studies carried out at Western Electric Co. near Chicago in the 1920ies. Two meanings frequently addressed in the literature are

- novelty, i.e. the notion that effects measured are confounded with (or simply induced by) the availability of a new, interesting experience or the increased attention paid to participant behaviours; and
- reactivity, i.e. the idea that effects are modified simply due to the fact that an evaluation study is conducted. This may lead, for example, to self-report biases or differential effects on the motivation to participate.

Furthermore, taking Hawthorne effects into consideration may also generate valid alternative explanations to causal inferences, and as such, this discussion refers to internal validity. It is likely that individual learners had not been exposed to much computer-based learning and/or CSSL courseware prior to the participation in this study. Due to novelty, therefore, the interest of participants working with the computer-based material may be unusually high (or unusually low) as compared to the interest of those more accustomed to using the computer for learning purposes. These interest levels may be associated with other traits possibly affecting situational judgement, such as sensitivity or readiness to help (in the evaluation and development of courseware). Then, it may not courseware use per se that effects learning gains and superior situational judgement but the confounding effects of novelty.

Associations between computer literacy (as measured by a 10-item sum score [K02_06s], see Appendix B.4, 2.6.) or computer familiarity (as extent of daily computer use [K02_05], see Appendix B.4, 2.5.) and courseware use and post-test situational judgement scores were explored. However, across all blended learning cohorts, no significant association was found between computer literacy or familiarity and courseware use (e.g. $r_{[K02\_06s]x[CUI_{TTL}]}$=.056, n=163, $CI_{95\%}$ from –.099 to .208; $r_{S[K02\_05]x[CUI_K]}$=.023, n=164, $CI_{95\%}$ from –.136 to .171). The same applies to the associations between computer literacy and post-test scores (e.g. $r_{S[K02\_06s]x[N\_MM\_S1R]}$=.070, n=151, $CI_{95\%}$ from –.096 to .223; $r_{S[K02\_05]x[N\_MM\_SRE]}$=–.088,

n=152, $CI_{95\%}$ from –.248 to .068). Thus, as far as novelty can be expressed in terms of prior computer use for learning purposes, it has not been found a confounding factor.

However, if – despite all concerns about reliability, validity and statistical power –these data can be taken to mean that there is an absence of relations between familiarity with computers and courseware use, the findings may further underscore the motivation of students to work with the courseware offered. Recent research findings indicate that familiarity with computers may also lead to shallow processing strategies which are less functional for learning (Wecker, Kohnle, & Fischer, 2007); this effect has not been found in this study. And, furthermore, a vast majority of students reported to have worked with the coursewares. These effects found may, however, limit the generality of the results; in this respect, they are essentially a matter of external validity.

One novel feature of this study pertains to the fact that courseware was newly implemented into the curriculum. Therefore, the findings may not be generalizable to future cohorts, and may be expected to wear off across time. And second, the effects found may be an artefact due to the fact that learners were expected to self-evaluate, to frequently observe their learning and to report progress and outcome. Therefore, the findings may not be generalizable to cohorts with less managed evaluation.

Some findings reported above, however, directly contradict both of these hypotheses. The last cohort evaluated in 2005/06 possibly comes closest to "normal operations"; here, evaluation was limited to an essential minimum, and programme management had largely handed over administration and maintenance to student assistants. And, much in line with the novelty and scale of evaluation hypotheses, courseware use variability was found much higher. By contrast, augmented increases and superior levels of situational judgement scores were found for high courseware use groups in this cohort, which – in scope and extent – resemble those of previous cohorts. Moreover, the highest between-group differences in self-efficacy level change were found for this 2005/06 cohort.

The *PROCON* curriculum itself is also a dynamic, innovative programme, especially with its pioneering attempts to introduce skill-based training to a (largely knowledge-based) higher education environment and to use computer-supported blended learning arrangements for this purpose. It may be hypothesized, therefore, that the findings of this study may not be fully generalizable to educational environments where skill-based trainings and blended learning scenarios are customary and generic. The same generally applies to the pedagogy and

educational approach as described in section 3.2. Various factors may be considered to play a major role in this respect, such as, for example, the amount of learner's personal control, their experience of personally meaningful skill acquisition, and the prospect of skills being acquired and certified with relevance for many psychology-related professions. It may also be true that the introduction of legal and formal rules for education and responsibilities of clinical psychotherapists in Germany in the late 1990ies has led many students seek alternative professional career paths; this may have contributed to an unexplored underlying motivation of the students.

The external validity of the findings may also be influenced by reactivity effects due to the study or evaluation itself. Participating students were also required to partake in pre- and post-tests and to submit post-module online surveys. Despite the fact that the pre-test was not used for selection purposes, the introduction of a pre-test and application procedure could have had deterrent effects and contributed to a sort of (self-)selection bias. Moreover, it is also a common phenomenon in educational research that testing and surveying affects learning; thus, reactive effects of the post-test and the module online surveys on both learning and outcome may be expected. In this study, students were asked not to prepare or review material simply for purposes of the post-test, and it had been communicated that certification was based on participation only, not on test-results. Of all 214 students taking the post-test, only 66 (30.8%) reported to having prepared (in any way) nevertheless, the half of which were tested in 2002/03 (i.e. the no-courseware cohort). A majority of the 66 preparers also reported to having spent less than 2 hours for preparation. Thus, the impact on learning and the interaction between testing and courseware use may therefore be considered limited. Moreover, in most educational contexts, assessment and evaluation are part of the learning culture. Thus, introducing research settings similar to the ones used in this study does not pose a major challenge, and the effects then be expected to be generalizable.

DIGEST 5.3

Potential threats to the theoretical, internal, and external validity of the inference that preparatory CSSL use positively influences gains and resulting ability to judge inter-personal situations and behaviours were discussed.

As evaluation measures used in this study had to be created from scratch, further inquiry into their psychometric properties is an issue for subsequent research. However, in the process of test construction and application, measures were taken to limit the potentially critical challenges of content validity and instrument change. For the time being, certain degrees of face validity is generally sustained.

A variety of threats to the internal validity of the inference that superior learning gains and resulting situational judgement levels may be attributed to the introduction and use of complementary CSSL courseware are discussed. No significant repeated testing effects were found. Confounding influences of biases found at pre-test could not be recognized in this study. Potentially confounding effects of training indicate that it may a worthwhile undertaking to further explore the relationships between conceptual learning by use of written materials, CSSL courseware, and hands-on training.

Possible novelty and reactive effects were discussed that may limit the generalizability or external validity of the findings. In particular, participants' use of the courseware as well as the intensity of curriculum management and evaluation may be regarded as unusually high.

5.4 Further Implications for Theory, Research and Practice

In the foregoing discussion of the validity of instruments and inferences, a number of suggestions for further research have already been put forward. For the most part, this subsequent discussion of further implications implicitly continues to address the issue of external validity and is guided by two main questions. Firstly, which of the factors studied, findings reported and conclusions drawn pertain to aspects of psychological and educational research and theories and frameworks outlined in sections 1 and 2 of this thesis? And, secondly, what assumptions can be derived that may help guide current and future implementation of Computer-supported Social Learning (CSSL) for soft skill development in higher education?

Due to the sheer amount of variables and data gathered across time, it is impossible to extensively discuss every aspect and each likely implication for theory, research, and CSSL practice. The following discussion, therefore, is restricted to a selection of implications and suggestions that might be of some interest to educational researchers, practitioners, and curriculum designers in institutions of postsecondary education.

CSSL Design and Integration.

As outlined in sections 1 and 2 of this thesis, social skills in general can be thought of as discipline independent, meaning that they are likely to be of importance in many occupations and professional domains. By contrast, the main CSSL courseware studied here centred around a specific set of procedural, stage-wise strategies used in group-based mediation. In addition, CSSL courseware is only one of many technologies that may likely shape future learning in postsecondary settings. For practitioners, three main questions arise out of this situation. Firstly, what idiosyncratic features of the CSSL courseware studied are likely to be varied if courseware is produced for other competencies or soft skills domains? Secondly, how may CSSL courseware be combined with other face-to-face and technology-supported elements to further enhance learning? And, finally, how can production costs be limited so that teaching and learning is cost-effective?

The impact of mediation-related courseware was chiefly studied here, relating to typical situational demands third parties face and prototypical slips they may avoid when guiding the parties in conflict though the stages of mediation. This means that the underlying conceptual model illustrated is of strategic and stage-wise nature. Stage-wise and/or strategic conceptual

models have also been developed for other areas in which socio-communicative competence plays a significant role, as, for example, in sales presentations (cf. Hershey, 2005), facilitation (e.g. McFadzean & Nelson, 1998), or problem solving (e.g. Spitzer & Evans, 1997).

By contrast, much inter-personal communication development draws on structural and process models to be applied to a multitude of inter-personal situations, roles, and perspectives (Schulz von Thun, 2003, 2004b). Moreover, both cross-situational Attitudinal approaches (such as *Principled Negotiation,* Fisher et al., 1991; 1998) and situation-specific techniques (such as *brainstorming*) are also often imparted. Thus, sequential stage models may probably be more the exception rather than the rule. When viewed from this perspective, depicting an on-going background narrative, divided into sequential segments may not always be possible, and two main questions are of interest to the findings reported, namely:

(a) What effects does (extreme) fragmentation of the background story have on retention and learning?

(b) What effects does the use of unrelated, individual video-scenes for the suggested task sequence (see Table 12 in section 3.3) have on retention and learning?

Whereas the former question is, at present, largely unexplored, first empirical findings (e.g. Bielecke, 2005) to the latter indicate that the learning gains achievable may be similar to the ones reported for successive tasks. Due to the higher number of different actors and the higher frequency of assembly/disassembly of the film-set, production costs for self-contained sequences may be somewhat higher than for serially shot video scenes. However, the fine 'granulation' of the self-contained material may allow for an integration and use in a variety of coursewares – which is not just economical but also worth pursuing from an instructional point of view when cross-functional and –situational socio-communicative competencies are the target of developmental endeavours. Here, a variety of self-contained multimedia-based sequences may illustrate relevant situations, contexts, and behaviours which, in both quantity and quality, are higher than those than can possibly be experienced by learners in simulated or live settings.

Recently, the web-based patterns and participatory technologies of what has been termed "Web 2.0" (O'Reilly, 2005) as well as the availability of broad-band internet access to most learners have given thrust to Computer-Supported Collaborative Learning (CSCL), one major paradigm of computer-supported e-learning (Koschmann, 2001). It may therefore also be worthwhile to study the effects of collaboration in software production and use. For example, behaviour generation and selection tasks could be collaboratively be reflected upon (e.g. Baker

& Lund, 1997), and both peers and subject matter experts could discuss and comment on situational features, behavioural options and models as well as accumulating further knowledge and conceptual models helpful for competence development. Moreover, pre-produced video scenes may also serve as templates for learner-based generation or production of videos (e.g. Kearney & Schuck, 2005; Derry, Hmelo-Silver, Nagarajan, Chernobilsky, & Beitzel, 2006).

This may also be another way of limiting the cost of CSSL production which possibly currently poses the greatest challenge to courseware designers. While multimedia production has become quite manageable, current learning management and/or content development systems do not yet support user tracking, and task, input and retrieval processes as exemplified, for example, in Table 12 (i.e. the task sequence, see section 3.3). Thus, on the one hand, specialist (and, therefore, typically expensive) knowledge and skills are still necessary to produce, embed, and deliver CSSL courseware which usually exceed those of most teachers and faculty in institutions of pre- and postsecondary learning, and, indeed, those of most committed to developing human resources and social skills. On the other hand, while CSSL coursewares may be expected to work best when restricted to specific behavioural models and/or a particular set of situations and (professional) roles, very limited markets and/or scopes of use may be assumed which renders CSSL production uneconomical for most areas. Currently, therefore, more promising areas of CSSL are more basic cross-discipline and/or cross-professional social skills.

In sum, while further evaluations of the cost-effectiveness in postsecondary educational settings are warranted, instructors and programme managers are encouraged to experiment with and report effects of variations of the task sequences and the blended learning mix.

CSSL in Postsecondary Education.

For the reasons outlined in section 5.3 it may not be safe to assume that the findings generalize to the population of all students in postsecondary education. While the background of the sample participants is relatively varied, it may not reflect that of any discipline, age, and skills group. One major factor likely to interact with courseware use is motivation. As outlined above, this study is based on volunteer participants who equated mediation skills acquisition with the development of competencies that are both relevant for professional tasks and personally meaningful. By contrast, many courses are not elective in postsecondary education, however, and, thus, motivation may become a crucial factor. Possibly, when novelty and other

related effects can be thought of as 'worn off', that is, when technologies are regularly used by students and/or staff in tertiary educational contexts (cf. e.g. Madigan, 2006), preparation with CSSL-based courseware can be thought of as equivalent to other forms of preparation currently widely used, such as reading. Then, far from being accepted with alacrity, this form of preparation (or follow-up) of face-to-face training is due to be met with the same rate of reluctance, possibly despite rising levels of commonplace information and computer literacy. For suggested reading materials, rates of adoption by learners – or, if put in transmissional terminology, "compliance with reading assignments" – may be considered to be rather low (Burchfield and Sappington, 2000), or, as Hobson (2004) put it: "Faculty face the stark and depressing challenge of facilitating learning when over 70% of the students will not have read assigned course readings" (p.1). Thus, ideas developed to foster other preparational and follow-up activities (such as reading) may also be translated to CSSL blended learning environments. This may be an especially fruitful approach as CSSL-based courseware may require prior conceptual scaffolding by means of reading materials and/or cooperative and collaborative learning units. Thus, many ideas for improving what is often referred to as "reading compliance" can be adapted for guiding practitioners implementing CSSL coursewares.

Additional face-to-face or classroom-based sessions may be considered the most viable approach to induce motivation to (a) acquire or develop the skills under view in general and (b) to prepare themselves by use of reading materials and courseware. In this project, linkages between textbook- or courseware-based preparation and the peer-tutored trainings were not systematically initiated, and, therefore, the actual degree of linkage reported by participants on a 6-point scale ([K03_02], see Appendix B.4, 3.) varies considerably between participants of the same training course (average range=2.36), between training courses (minimum $M_{[K03\_02]}$=1.64, maximum $M_{[K03\_02]}$=3.5, average $M_{[K03\_02]}$=2.72), and between cohorts (in 2003, cross-course $M_{[K03\_02]}$=2.80; in 2004 $M_{[K03\_02]}$=2.49). It may be worthwhile, therefore, to further and systematically integrate preparational materials into classroom sessions, to discuss the relevance of reading materials as well as CSSL coursewares prior to delivery or "assignment", and to allocate in-class time to scaffolding by previewing recommended materials (Hobson, 2004).

Moreover, external incentives for the use of preparational materials may also be introduced. Control mechanisms may also serve self- or peer-monitoring, and thus, the purposes of learners themselves, and the group of learners in which learning is embedded.

However, caution is well indicated as many ideas implicitly subscribe to transmissional educational approaches that aim at increasing the levels of enforcement, control, or coercive power. These include threatening with test, random questioning (e.g. McDougall & Cordeiro, 1993), restricting course admission or continued participation, and classroom assessment. Computer technologies potentially allow for even higher degrees of control; for example, log-files may be used to examine the actual use of preparational materials by students.

Assessment of Socio-communicative Competence.

For various reasons, "authentic" assessment and assessment of competencies and skills (rather than knowledge) is likely to become indispensable in postsecondary educational contexts and institutions of higher education which also serve as certifying bodies. Currently, it remains true that (high-stakes) assessment of complex skills such as socio-communicative competence is far from being well-engineered, psychometrically and technically mature.

In this study, situational judgement was primarily measured as knowledge-based as students were asked to identify with the authors of the conceptual knowledge (as subject matter experts) and rate the behavioural alternatives offered in terms of their perspectives. This approach may serve as a template guiding the future development of assessment batteries for less subjective skill level estimation.

Interestingly, however, similar outcomes were obtained for scoring based on the opinion of subject matter experts and for scoring based on empirical keying, meaning that scores were awarded according to the percentage of respondents. This finding is much in line with recent research (Krokos et al., 2004; Lievens, 2000; MacCann, Roberts, Matthews, & Zeidner, 2004; Morath, Curtin, Brownstein, & Christopher, n.d.). However, it might also raise two additional issues both of which pertain to the models of assessment (see Figure 4, section 2.2) and development (see Figure 8, section 2.4) of socio-communicative competence elaborated upon.

First, CSSL courseware may help learners to align their judgement of situational demand characteristics and the behavioural responses of a communicant to those of subject matter experts. However, little is known as to the degree of congruence between their overt rating and their covert cognitive-affective and attitudinal processes. It is an interesting question, therefore, to what extent differences remain, in any given situation, between the learners' own personal appraisals of the situation, their internal ambivalences, and their overt response.

And, second, it is an interesting question, if the CSSL-based software actually contributes to behavioural change. In accordance with the developmental model (see Figure 8, section 2.4), modelling and refection processes should eventually result in behavioural changes. One method to access these in written situational judgement tests is to solicit written responses before presenting the behavioural alternatives for rating; as such, they can be thought of as a standardized form of video-based situational interviews (Latham et al., 1980), as, for example, used in the post-tests in this study (see Appendix C, Section III; analysis is still pending). However, while situational interviews (i.e. what people say they would or will do) are generally found to predict actual performance (i.e. what people actually do), correlations found between these two are typically rather low (McDaniel, Whetzel, Schmidt, & Maurer, 1994).

Thus, situational judgement tests may be considered insufficient for assessing the full impact of Computer-Supported Social Learning. In line with the approaches to the assessment of socio-communicative competence (see Figure 4, section 2.2), complementary assessment strategies may be drawn on to implement a multi-method-approach to assessment that can provide additional insights into the potentials of CSSL. Actual or performance assessment strategies may help answering the question to what extent the behavioural models presented in the coursewares were actually adopted, i.e. not simply as knowledge but as employable behavioural options or performable skills. Secondly, strategies aimed at tapping inner processes – such as thinking-aloud or methods drawing on the 'Inner Team' (Schulz von Thun, 1998, 2004b; Schulz von Thun & Stegemann, 2004) metaphor – can complement learning impact evidence. And, finally, both research and educational practice could also aim at determining the value of derivative work samples (such as journals, essays, collaboration reports or other written materials) in learners' portfolios for purposes of an assessment of socio-communicative competencies.

Theory-based Research Agendas.

Recently researchers have begun to argue that the introduction of interactive learning technologies possibly serve not all learners to the same extent (e.g. Gibson, 2001; Baldwin & Sabry, 2003). In principle, this argument may also apply to the CSSL coursewares used in this study as they generally employ the paradigmatic sequence detailed in Table 12 (see section 3.3) which may be viewed as making unequal use of the various paths to Social Learning. In terms of Kolb's styles of learning as depicted in Figure 6 (see section 2.4), preparational use of CSSL courseware may be hypothesised to be especially suited to learners preferring concrete-

experiential (CE) and reflective-observational (RO) styles of learning. By contrast, however, it may be least appropriate for those preferring active experimentation (AE).

All 40 participants of the 2005/06 cohort partaking the post-test were asked to complete Kamentz' (2006) Learning Style Inventory, a 20-item questionnaire employing forced-choice alternatives between statements based on preferential styles of learning according to Kolb, adapted to higher education environments. For each choice of learning style, a point was awarded, and learners were categorized according to the style with the highest absolute score. One participant was excluded from the analyses due to equal scores across styles; for all remaining participants, the most preferred style of learning and self-reported courseware use is detailed in Table 42.

Table 42

CSSL Courseware Use and Most Preferred Style of Learning (2005/06 Sample Base)

Self-reported Courseware Use[a]	Style of Learning: Abstract Conceptuali-sation	Active Experi-mentation	Concrete Experience	Reflective Observation	Σ n
0-30%	0	4	2	4	**10**
40-70%	3	2	3	1	**9**
80-90%	1	1	3	1	**6**
100%	1	1	6	6	**14**
Σ n	**5**	**8**	**14**	**12**	**39**

Notes. Number of participants (n) given. One participant was excluded due to the same score across all learning styles. [a]Based on self-reported Conflict Resolution in Groups courseware use [K02_023, %] Grouping Alternative B in Tables 26/27 (see section 4.2).

Interestingly, the results indicate that there may be some biases ($\chi^2(9)=11.49$, C=.477, p=.244) in the hypothesized direction. However, the sample is too small and the instruments possibly lack validity, so the issue is best left to future investigation.

Future research may also explore other links between CSSL and its theoretical bases to provide more evidence as to the underlying theoretical assumptions of computer-supported social learning. For example, behavioural models shown in CSSL-based courseware that only had a near-zero chance of occurrence prior to software use should be adopted by users, even without subsequent training.

Another worthwhile line of research might explore the differences between conceptual learning and by use of books or written materials and CSSL which can also be thought of as two variants of situating conceptual learning. According to theory, there should be major differences between situating learning by means of high-fidelity (e.g. video-based) and low-fidelity (e.g. text based) contextual accounts. Research in this area could follow similar paths as

item stem and response analysis in situational judgement testing (Funke & Schuler, 1998; Havighurst et al., 2003; Motowidlo, Dunnette, & Carter, 1990).

It also remains an interesting question yet what effect CSSL has on learner's self efficacy – and its accuracy. As outlined above, two opposite trends may be assumed. Gains are likely where learners develop and test their practical skills; on the other hand, CSSL users may view expert models and sense that they would not (yet) be able to reproduce the model's behaviours and/or that they lack experience and expertise possibly needed to attain equivalent skill levels. This may, in turn, reduce positive biases of self-image and, thus, reduce self-efficacy.

And, finally, CSSL affordances and learning opportunities in various other settings remain to be explored such as the effects of use after training, the influence of motivation and culture, or the interplay between training and CSSL courseware use.

DIGEST 5.4

In addition to the suggestions for future research put forward in the foregoing, a selection of questions and assumptions are discussed that pertain to this study's underlying theories of learning, competence development and assessment as well future forms of Computer-supported Social Learning (CSSL) in postsecondary educational contexts.

Some idiosyncratic features of the CSSL courseware studied limit generality as they can be assumed to interact with the independent variable. Of these, some are likely to be varied if courseware is produced for other competencies or soft skills domains. Among these features is an underlying conceptual model which is sequential stage model, allowing for an on-going background narrative, and the high degree of participant motivation. Moreover, learners with some preferred styles of learning may also be more hesitant to use CSSL courseware.

Moreover, the links between CSSL courseware and other forms of preparatory and supplementary individual learning as well as face-to-face learning need to be explored and experimented with to better understand the scope and limits of the underlying theories and to enhance learning outcome and the level and accuracy of related self-efficacy. It is also suggested that CSSL may also be combined with other face-to-face and technology-supported elements to enhance collaborative learning as well as production as a means of limiting production costs.

And, finally, it is also an interesting question for both research and practice to further explore the capabilities and affordances of CSSL in applied contexts. As institutions of postsecondary education are also certifying bodies, there is a need to develop skill-based forms of assessment, for which this study may serve as a template in that learners are asked to judge situations based on their knowledge. Complementary forms of assessment may explore differences between their knowledge-based situational judgment and learners' personal appraisals of the situation, their internal ambivalences, and their own behavioural responses.

5.5 Summary and Concluding Remarks

In postsecondary educational contexts, learning environments have traditionally been geared towards the construction, generation, refinement, assessment, and certification of individual knowledge. By contrast, there is a growing recognition that graduates from institutions of higher education may need to be equipped with higher levels of employability and transferable, cross-disciplinary and cross-professional (certified) competencies. One of these is socio-communicative competence, or the social skills of adult learners pertaining to a specific set of situations, roles and/or professional functions.

Classroom-based training, hands-on simulation, and personal coaching and supervision incorporating and making use of the opportunities for inter-personal communication, the social relations and the contingencies of the learning culture between those present have traditionally been viewed as epitomes of planned social skills development. Recent innovations in what is usually referred to as advanced learning technologies have provided opportunities for changes of traditional delivery modes. Current mainstream research and practice in this area explores approaches to computer-supported collaborative learning (CSCL) one premise of which is an extant system of learners.

By drawing on models for socio-communicative competence development and assessment, as well as Bandura's Social Learning Theory as a meta-theoretical base, an alternative theoretical framework for the expedient use of computer technologies for planned social skill development has been put forward in this study and referred to as *computer supported social learning* (CSSL).

CSSL essentially focuses on individual learners and aims at scaffolding conceptual knowledge or theories of action (usually explicated in books or lecturettes) assumed to be helpful for mastering social situations addressed by the concepts or theories. For this purpose, individual learners are typically provided with a number of high-fidelity presentations of complex socio-communicative situations by means of digital media, asked to actively generate responses to these situations, observe and rate alternative responses of the communicants, which elicits additional feed-back by subject matter experts. Additional reflection-oriented tasks may help deepen and elaborate the conceptual representations of action, situation, and/or personal characteristics involved.

In this study, the impacts of the introduction and use of tailored, interactive CSSL-based coursewares for mediation training were evaluated in a university-based setting between 2002 and 2006. Blending traditionally delivered training with supplementary use of the coursewares for preparational purposes was expected to effect augmented increases and higher resulting levels of socio-communicative competence required by mediating third parties in group-based conflict-resolution.

Across cohorts, 272 different students took part in one or more programme modules of which 195 participants completing the facilitation, problem solving and mediation training curriculum within one year and cohort. An overall integrative evaluation strategy was followed to take account of the various characteristics and restrictions often found in tertiary educational fields. Both the multiplicity of module participant groups within a one-year-cohort, the four entire consecutive student cohorts with differing degrees of computer support as well and ex-post facto groups were identified as most proximate to quasi controls.

Data reported in this study were obtained in pre-tests, in post-module online surveys, by means of log files, and post-tests, the latter of which comprised one of two parallel versions of a video-based situational judgement test (SJT) as this seemed an apt and effective way to assess mediation-related socio-communicative competence in larger samples which may also serve as a more general template for institutions of postsecondary education in need to develop skill-based forms of assessment.

As evaluation measures used in this study had to be created from scratch, ordinal levels of measurement an non-normality were assumed for the majority of data obtained (unless reasonably established). To avoid problems of statistical conclusion validity as well as to ease interpretation and justification, statistical procedures were selected and adapted accordingly.

Across cohorts, significant medium associations were found between self-reported courseware use and SJT 'pick-best' scores ($r \approx +.4$). Substantial post-test differences in situational judgment were found between 'low' and 'medium' (or 'partial') use groups with the latter consistently displaying superior levels (Cohen's $d \approx +.8$ to $+1$). The differences seem to be strongest where CSSL courseware was used as a complementary, not substitutional means of preparation. Augmented score gains were found in one cohort which can be regarded as a medium effect ($d \approx +.5$).

No consistent significant associations were found between courseware use and post-test levels of conceptual knowledge or interest in subject matter. Self-efficacy gains and post-curricular levels were found either equal or higher for the "high courseware use" ex-post facto groups, but the found effects are rather small (*d* typically between +.0 to +.3), possibly due to measurement levels, lacks of reliability and diminished statistical power of the instruments used.

Potential threats to the statistical, theoretical, and internal validity of the inferences were considered. While no significant effects of repeated testing nor confounding influences of biases found at pre-test could be recognized, future research may further investigate the psychometric properties of the instruments and may rule out remaining threats to the internal validity of the findings. Possible novelty and reactive effects as well as distinct features of the sample and setting such as high levels of participants' motivation and the intensity of curriculum management and evaluation were also discussed.

Features of the CSSL courseware, its combination with other forms of learning as well as production budgets are likely to be varied if courseware is produced for other soft skills domains and applied settings. It is also an interesting question if the findings can be reproduced with other forms of assessment methods. In the concluding section, therefore, selected implications for theory, research, and practice were explored.