Running head: Modelling self-regulating and self-directing constructs

Modelling Self-Regulating Capabilities and Self-Directing Capabilities, of Adult Students: Relations with Learning Outcomes and Labour Market Success

Monique Bijker, Marcel van der Klink, and Els Boshuizen

Centre for Learning Sciences and Technologies (CELSTEC)

*Correspondence concerning this article should be addressed to Monique Bijker, Open University of the Netherlands, Centre for the Learning Sciences and Technologies, P.O. Box 2960, 6401 DL, Heerlen, the Netherlands. T: +31 45 5762814; E: Monique.bijker@ou.nl

Modelling self-regulating and self-directing constructs

Abstract

This study investigated generic competences predicting students' learning outcomes and labour market performance. The literature revealed four predictive self-reported concepts: self-efficacy, self-regulating learning capabilities, self-directing learning capabilities, and self-directing career capabilities. However, there was a lack of conceptually clarity, in combination with a lack of empirical evidence on how the concepts were related. Hence we departed from the four respective theoretical frameworks to safeguard the content validity, constructed theory-based items and scientific measures, using the Rasch model, and modelled the constructs in a structural equation model. Our models revealed a dynamic intra-individual causal system in which self-efficacy and self-regulating learning capabilities are predictors of self-directing learning capabilities that operate as a mediator towards grade point average. Self-efficacy acts as a significant direct predictor of achieved European Credit Transfer System credits (ECTS). Self-directing career capabilities showed to be two-dimensional.

Keywords: self-regulating capabilities; self-directing learning capabilities; self-directing career capabilities; Rasch rating scale model; structural equation modelling

Introduction

Rationale for the Study

Conceptual confusion.

This study focuses on the development, and validation of theory-based measurement instruments to model self-efficacy, self-regulating learning capabilities, self-directing learning capabilities, self-directing career capabilities, and educational outcomes. The rationale for improving and developing instruments to measure and model these capabilities is that the literature considers them as important trainable generic capabilities, and predictors of study success and/or labour market outcomes. Generic capabilities, in contrast with domain-specific capabilities, are viewed as transferable to different situations and contexts. Generic capabilities are supposed to facilitate future competence development in the professions, in the workplace, and in follow-up training. Hence they are supposed to enhance career opportunities and employability.

From a conceptual perspective it is important to investigate whether these capabilities are distinct concepts; from an empirical perspective it is relevant to identify whether they predict different outcomes. To date, these generic capabilities have been theorized in many different ways in multiple disciplines, such as educational psychology, andragogy, and organisational psychology. Corresponding theoretical frameworks can be identified in social-cognitive theory, the motivated strategies for learning theory, adult learning theory, and career theory. Some theorists such as Pilling-Cormick and Garrison, and Jossberger, Brand-Gruwel, Boshuizen, and Van de Wiel (in press) attempt to theorize on possible differences between the self-regulating and self-directing concepts. However, empirical evidence concerning the distinction, overlap, or relations between these different generic capabilities is still lacking, just like evidence that informs us on their possible differential predictive power towards distinct learning outcomes in different domains or professions.

From a curriculum design perspective it is important to identify which of these theorized, and sometimes operationalized capabilities are the most powerful predictors of learning outcomes, and/or labour market success in transition stages from school to work, or from work to school, because such information has important implications for curriculum and training design.

Measurement and modelling issues.

In addition to conceptualization issues, we focus on developing operationalizations of self-regulating and self-directing concepts in psychometrically valid and accurate measures, which can only be realized by aligning the three levels of theories (conceptual, methodological, and statistical theories; Mellenbergh, 1980, 1996). For our generic capabilities this concerns three research steps.

First, a non exhaustive review of the concepts and their associated theoretical frameworks is carried out to address the first research question "Can generic self-regulating capabilities and self-directing capabilities be identified as conceptually distinct capabilities?" The conceptual review will function as the point of departure for the content validity of operationalizations of the four generic capabilities.

Second, following Bond and Fox (2007), we will argue and demonstrate that good measurement leads to profound theory, and unbiased results. Our operationalized generic capabilities will be modelled in the Rasch rating scale model, and transformed into interval measures, before statistic techniques are carried out to identify the correlational relations between the capabilities, and before causal modelling will occur. This second research step will help us in addressing the second research question "Do the self-regulating and self-directing constructs fit the Rasch rating scale model, to produce Rasch interval measures?"

Eventually, our third and central research question can be addressed: "Can the self-regulating and self-directing learning capabilities (expressed in Rasch interval measures) be combined in a structural model that predicts different learning outcomes?"

In the next section the conceptual frameworks underlying the generic capabilities will be summarized. Social-cognitive theory will pass the revue as the fundamental for the motivated, self-regulating learning concepts such as self-efficacy and self-regulating learning capabilities. The next framework is adult learning theory, from which the concept self-directing learning capabilities has emerged, followed by career theory that underlies self-directing career capabilities. This section may assist us in addressing the first research question.

Conceptual Perspectives

Self-efficacy.

Bandura (1986) can be considered as the founder of social-cognitive theory. In his theory Bandura modelled expectancy concepts (self-efficacy, and outcome expectations) together with values (task values) towards learning outcomes. Specifically self-efficacy has shown to be an important predictor of learning outcomes. Self-efficacy is an expectancy concept that is defined as "People's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986, p. 391). Self-efficacy is distinct from self-confidence, which reflects more general competence beliefs (e.g. "I am good at maths"), while self-efficacy refers to more specific and situational judgments of capabilities (i.e. "I am confident that I can successfully accomplish this research assignment"; Pintrich & Schunk, 2002). Social cognitive theorists suggest that it is necessary to differentiate between very general self-concept beliefs and more specific judgments about domain-specific tasks. Task-specificity and correspondence with the outcome of interest optimize the predictive power towards academic outcomes (Marsh, Roche, Pajares, & Miller, 1997). It implicates that self-efficacy judgments for specific tasks vary as a function of individual and contextual differences (Pintrich & Schunk, 2002). Self-efficacy is also very different from self-worth or self-esteem, which concerns the individual's affective evaluation of the self (Eccles, Wigfield, Flanagan, Miller, Reuman, & Yee, 1989; Pintrich & Schunk, 2002). In other words, the feelings we have about ourselves. In comparison with self-confidence and self-esteem self-efficacy is the most contextually defined cognitive self- concept (Pintrich & Schunk, 2002). Self-efficacy has proven to be an important predictor of achievement behaviour (Linnenbrink & Pintrich, 2002, 2003).

Self-regulating learning capabilities.

The merit of Pintrich, Smith, Garcia, and McKeachie (1991, 1993) is that they have extended Bandura's model by hypothesizing that cognitive and metacognitive learning strategies would mediate the relations between motivations and expectancies and learning outcomes. Their research has revealed that to a large extent self-efficacy defines behavioural engagement in academic learning (Linnenbrink & Pintrich, 2003).

Pintrich et al.'s (1991, 1993) "Motivated Self-regulated Learning Theory, emphasizes the importance of self-regulation. Self-regulation is defined as metacognitive planning, monitoring, directing and evaluating the cognitive strategies that are used (Pintrich et al., 1991, 1993). Pintrich (2004) elaborated self-regulation as a dynamic, interacting intra-individual system of motivations and learning strategies. He conceptualized self-regulation as an open intra-individual system, within which motivations, expectancies, and learning strategies continuously interact with each other, and in which all variables reflect a certain degree of self-regulation. The intra-individual systems interacts with external systems, such as learning tasks, the curriculum, and for example work environments. Within the intra-individual system of interacting motivational and learning strategy variables, self-regulating capabilities (SRLC) are conceptualized as effort, focus, persistence and perseverance in academic learning (Linnenbrink & Pintrich, 2003). Self-regulating capabilities act like a spider in the web of the intra-individual self-regulating system. Self-regulating capabilities play a crucial role in metacognitively regulating and directing the motivations and cognitive strategies towards the learning outcomes. A study of Bijker, Wynants, and Van Buuren (2006), has validated the intra-individual system theory of Pintrich (2004) in the context of a redesigned research curriculum for psychology students at the Open University of the Netherlands. Bijker et al., and Van Buuren (2008) indicate that self-regulating capabilities act as a mediator between motivations such as intrinsic task value, instrumental task value, self-efficacy, and/or test anxiety, and learning outcomes in complex tasks, such as psychological research tasks.

Pintrich et al. (1991, 1993) allowed researchers to select the variables of interest from their self-regulating system. Most scholars in this stream of thought consider the single concept self-regulating learning capabilities as self-regulating capabilities. In our current study we will focus on this single metacognitive concept and refer to it as SRLC.

Self-directing learning capabilities.

Adult learning theories: a tower of Babel?

The concept of self-directing learning has emerged from adult learning theories (Hiemstra, 2000; Knowles, 1975). Originally, adult learning theories have also defined self-directed learning in terms of planning, monitoring, and evaluation (Hiemstra, 2000). The ever expanding and confusing conceptualizations of self-directed learning may have obscured the key-characteristic of the concept, namely the individual difference concept 'independent learning', situated in informal learning situations. Knowles (1970) described self-directed learners as "individuals [that] take the initiative, with or without the help of others, in diagnosing their learning needs, formulating their learning goals, identifying human and material resources for learning, choosing and implementing appropriate learning strategies, and evaluating learning outcomes" (Knowles, 1975, p.18).

Over time, the individual difference concept self-directed learning has been influenced by policy, economic, and socially desirable goals such as lifelong learning (macro perspectives; Nieuwenhuis, Gielen, & Nijman, 2008), transforming the concept into a highly trainable concept, and eventually, an instructional method (Guglielmino, Long, & Hiemstra, 2004, p. 7) that is generalized to formal education situations in lower schools, colleges and universities (Merriam and Caffarella, 1999, as cited by Guglielmino et al., p. 9).

To date there is little evidence of self-directed learning as a process (Owen, 2002). Descriptive studies in informal learning contexts have dominated the field. The most popular instrument to measure and investigate the concept quantitatively (Guglielmino, 1977) has been criticized for its construct validity, and the likelihood that the instrument is measuring multiple dimensions simultaneously. In essence, adult learning theories capitalize on a self-initiated or maturated kind of self-direction in learning that is supposed to be characteristic to a certain degree for all adults, whether they are high-qualified or low-qualified (Raemdonck, 2006).

To avoid the confusion of tongues in the body of predominantly qualitative studies in informal learning contexts, and the dubious quality of the prevailing measurement instruments that are used in the relatively scarce quantitative self-directed learning research , this study focuses on the original conceptualization of self-directed learning in terms of independent learning, planning, monitoring, and evaluation, and the current work of Raemdonck (2006) who has resituated self-directed learning in such a way that it unites strategic learning behaviour in formal and informal learning. Hence, we relabel the self-directed learning concept in self-directing learning capabilities (SDLC), acknowledging its expertise-related, formal and informal nature. Raemdonck (2006) provides a contextual description for SDLC as follows: "[...] a characteristic adaptation to influence work-related learning processes in order to cope for oneself on the labour market" (p.13). She describes series of informal and formal work-related learning activities that result in the achievement of learning-related goals, such as mastering new tasks or updating skills and knowledge. Raemdonck emphasizes adaptiveness, influence, and coping. Her conceptualization is relevant for transition situations from school to work (or vice versa), and for occupational and professional education and practice.

Raemdonck (2006) has used advanced methods to develop theory-based operationalizations for SDLC, such as confirmatory factor analysis (CFA), to enable quantitative research. However, her instruments are new and need further development. In addition to her informal research there also is a need for research that investigates how formal learning and education might affect SDLC.

Theoretical attempts to integrate self-regulating learning capabilities and self-directing learning capabilities.

Current educational theories strive for more conceptual clarity of concepts such as self-regulating learning capabilities and self-directing learning capabilities. Pilling-Cormick and Garrison (2007) have conceptualized self-regulating learning capabilities as the (covert) management of the cognitive/affective, internal learning environment (like SRLC in Pintrich's (2004) intra-individual system), and self-directing learning capabilities as the (overt) management of the external learning environment. Other theorists in the domain of preparatory vocational education, such as Jossberger, Brand-Gruwel, Boshuizen, and Van de Wiel (in press) have suggested that self-regulating learning capabilities are operational at the micro-level and self-directing learning capabilities at the macro-level, which implicates that they connect self-regulating learning capabilities with task execution, and self-directing learning capabilities with complete learning trajectories. They approach both concepts from a normative and developmental point of view. Pilling-Cormick and Garrison's (2007) and Jossberger et al.'s (in press) attempts to integrate self-regulated and self-directed learning stress the need to verify their propositions with theory-based operationalizations of both concepts. If Pilling-Cormick and Garrison's (2007) conceptualization is correct, we can hypothesize to identify two differential constructs in our analyses: self-regulated learning capabilities (SRLC) and self-directing learning capabilities (SDLC); if Jossberger et al.'s (in press) suggestions are correct, self-regulated learning capabilities might either be nested in the theory-based operationalization of self-directing learning capabilities (SDLC) or specifically reflect task execution features, while self-directing learning capabilities (SDLC) will specifically reflect longer term planning.

Self-directing career capabilities

Career Theory.

Raemdonck (2006) has illustrated that the concept self-directed career capabilities (SDC) has emerged from career theory, which has its fundamentals in multiple disciplines. Self-directed career capabilities (SDC) are defined as "a characteristic adaptation to influence career processes in order to cope for oneself on the labour market" (Raemdonck, 2006, p.13). Career research predominantly concerns the study of proactive career behaviors, career planning, career management development, consultation, and networking (Claes & Ruiz-Quintanilla, 1998). From the 1980s onward career theory has started to focus on the life span, while simultaneously, as a consequence of downsizing trends in organizations and the unpredictability of organizational life, introducing concepts such as the 'modern career' or the 'boundaryless career' (Mirvis & Hall, 1994), inspired by new views on the nature of employees' relationships with employing organizations. Career theorists have suggested that self-directedness in career processes should be an integral part of a modern psychological career contract. According to King (2004) "… negotiating or bargaining with an employer has become familiar in discussions of psychological contract negotiations, and is particularly relevant to careers that span organizational boundaries, where negotiation of terms and obligations may be frequent and explicit" (p. 121). In other words, a personal career should be the shared responsibility of individuals and organizations.

From the 1980s onward, concepts such as lifelong learning, mobility, and employability have emerged that have emphasized the individual's agency in managing personal career opportunities (cf. King, 2004; Semeijn, Van der Velden, Heijke, Van der Vleuten, & Boshuizen, 2005; Van der Heijden, 2002; Van der Klink & Boon, 2002). Underlying such concepts are economic and policy rationales on competitive advantages of organizations and markets, if employees and civilians have acquired increasingly more sophisticated competences, resulting in added value for organizations, and more favourable economic positions of national and international markets (Nieuwenhuis, Gielen, & Nijman, 2008).

Relations between self-regulating, self-directing capabilities, and outcome measures

Correlational studies (Pintrich & De Groot, 1990), and experimental studies, in which feedback was used to empower self-efficacy (Schunk, 1982 in Pintrich & Schunk, 2002, p. 10; Jackson, 2002) have provided evidence for the relation between self-efficacy and the use of cognitive and metacognitive strategies. Wolters, Pintrich and Karebenick (2003) have reported regression coefficients between self-efficacy and learning strategies in a range of .10 and .67 in middle and high schools (when other motivational constructs were controlled for).

Self-regulating learning capabilities have shown to act as significant predictors of learning outcomes in complex domains, with partial eta squares in ranges between .04 and .08 (cf. Bijker et al., 2006; Van Buuren, 2008). The whole system of self-regulated variables however, as supposed by Pintrich (2004), can explain fifty percent of the variance in learning outcomes, depending on the instructional design and expertise level of the groups under study. There is a rich body of evidence on self-regulated learning and SRLC, both in initial and post-initial education contexts (cf. Blom, Severiens, Broekkamp, & Hoek, 2005; Chen, 2002; Pintrich, 2000; Pintrich et al., 1993; Wild & Schiefele, 1993; Bijker et al., 2006; Van Buuren, 2008).

Thus far we know that both self-directing learning capabilities and self-directed career capabilities are determinants of employability (cf. Fouarge, De Grip, & Nelen, 2009; Parker, Hall, & Kram, 2008; Raemdonck, 2006). Raemdonck (2006) has found correlations between self-directing learning capabilities and self-directing career capabilities of $r = .71$ in a sample of low-qualified employees, and a correlation of $r = .55$ in another sample including both high- and low-qualified employees. In addition she has identified relations between self-directing learning capabilities and self-directing career capabilities towards employability.

King (2004) has theorized that self-efficacy is a determinant of self-directing career capabilities. One of the few studies that have revealed such a relation is a longitudinal study, in the context of vocational education, carried out by Pinquart et al. (2003). These researchers found that self-efficacy, in combination with better grades, was a significant predictor of being employed in the transition stage from school to work. Pinquart et al. also found that self-directing career capabilities mediated the relation between self-efficacy and employability. However, neither self-regulating learning capabilities, nor self-directing learning capabilities were researched in that study, and the operationalizations of the constructs were vague.

Conclusion conceptual perspectives.

Table 1 summarizes the major findings from our conceptual summary concerning each of the four concepts.

Table 1

Classification of Self-Regulating and Self-Directing Concepts, Descriptions, and Relations.

Classification	**Concept**	**Description**	**Relation with learning or labour market outcomes**
Bottom-up, social cognitive	Self-efficacy	The most contextually defined social-cognitive concept, and an expectancy variable in Pintrich et al.'s (1991, 1993) intra-individual system theory. Defined as "People's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances".	Is both a strong predictor of learning strategies and learning outcomes. Might also be a predictor of employment in transition stages.
	Self-regulating learning capabilities (SRLC)	Social cognitive. Both the overarching principle and central self-regulating learning capability in the intra-individual system theory. SRLC, the central metacognitive construct is operationalized in terms of effort, focus, persistence and perseverance.	Has shown to mediate relations between motivations and expectancies towards cognitive strategies. Predicts learning outcomes such as task performance and academic achievement.
Top-down; political,	Self-directing	Many different	Self-directing learning

economic, societal.	learning capabilities (SDLC)	definitions, in terms of individual differences; metacognition (planning, monitoring and evaluation); management of the external learning environment; coping mechanisms; developmental stages; education. Originally: independent learning in informal contexts.	capabilities, in terms of independent learning and planning, monitoring, and evaluation, have shown to predict employability.
	Self-directing career capabilities (SDC)	From a personal difference concept towards a developmental concept and interactional concept. Defined as "a characteristic adaptation to influence career processes in order to cope for oneself on the labour market". Studied and operationalized differently in multiple disciplines.	Self-directed career capabilities have shown to be related to employability.

Based on a non-exhaustive summary of theory-based literature we may infer that the self-regulating and self-directing concepts are clearly distinct classes of concepts. We can classify the self-regulating concepts as bottom-up concepts, while the self-directing concepts can be classified as top-down concepts.

Both self-efficacy and self-regulating learning capabilities are strongly rooted in social cognitive theories, and a rich body of empirical research. We are well informed about their relations with cognitive strategies, and outcome measures in formal contexts. Self-efficacy can also be related to informal contexts. Self-efficacy and self-regulating learning capabilities are well supported by validity studies (cf. Pintrich et al.,1993), and empirical studies in a broad range of formal learning contexts.

However, studies that investigate their impact in informal learning contexts, such as the workplace and the labour market, are relatively rare.

In contrast, both self-directing concepts seem to evolve from personal difference concepts towards developmental, interactional, and educational concepts, influenced by economic, political, and social developments, and normative goals. The concepts appear to translate dynamic macro-level developments in an emphasis on the individual's agency, the responsibility to learn continuously and independently, and to cope for oneself in daily life or in the labour market. Consequently, both self-directed concepts show problematic characteristics, reflecting chameleon like, multi-dimensional features, and tendencies to fit all kinds of theoretical frameworks and contexts. Macro-perspectives pervade the individual difference characteristics, and turbulent changes in macro-systems model the concepts in the same pace. The paradox regarding the self-directed learning concept is that the predominant qualitative research paradigm tends to underestimate the importance of accurate measures and quantitative research, while simultaneously allowing deterministic and normative policy and economic goals to affect the concept. In contrast with the self-regulating framework, the predominantly interpretative self-directed learning frameworks do not offer fine-grained measures to investigate the relations of the concept with other variables of interest, such as social cognitive concepts. Self-directed learning beliefs versus self-regulating learning capabilities, are rooted in inherently different methodological traditions (interpretative qualitative versus empirical quantitative). Since the top-down concepts continuously change their conceptualizations and operationalizations, their relations with other concepts and outcomes are uncertain and incomparable.

The tentative answer on our first research question "“Can generic self-regulating capabilities and self-directing capabilities be identified as conceptually distinct capabilities?” is that the two classes of concepts are inherently different in nature. The abundance of different conceptualizations and operationalizations of the top-down concepts makes it almost impossible to further investigate similarities and differences, relations between the concepts, and their predictive power towards relevant outcome measures.

Hence, the next step is using real data, and operationalizations that are grounded in a rigorous measurement theory, to verify whether four different constructs can be distinguished. A fundamental, scientific measurement theory will be applied that defines measures, and can construct person capability measures independent from the items, and items independent from the persons: the Rasch model (Bond & Fox, 2007; Linacre, 2009a).

In the next section we will first describe the methods that are used to gather and process the data. In the subsequent sections the practical and theoretical rationales underlying the Rasch model are described, followed by Rasch analyses to validate theory-based constructs.

Methods

Participants

Respondents were 449 adult students, predominantly of the Open University in the Netherlands who already possessed a higher education degree (a Higher Vocational Education bachelor degree, either or not an accomplished pre-master, a Bachelor of Science degree, or a Master of Science degree in a different domain). The percentage of respondents that was currently involved in the final stage of a Master of Science program was 48.9%. Respondents came from three different disciplines: educational sciences (51%), psychology (31%), and business administration (17.7%). Educational science students were predominantly teachers in primary, secondary, vocational, and higher education. Respondents in business administration came from three different business schools in the Netherlands and Flanders. The average age of the respondents was 39.90 (*SD* = 10.96). Most respondents were female (65.5%), versus 34.5% male.

Instruments

The points of departure for the composition of our questionnaire were the self-efficacy scale and SRLC scale of Pintrich et al. (1991, 1993), and the SDLC scale and SDC scale, as developed by Raemdonck. The 28 items as validated by Raemdonck, (14 for SDLC and 14 for SDC) were empowered with 33 extra items to increase the opportunity that every theorized component of both constructs would be included in the operationalizations of SDLC and SDC (see Table 2), which is important for both content and construct validity. As Messick (1995) points out: both construct underrepresentation and construct irrelevant variance are threats for construct validity. Some self-efficacy items were added to fit the domains of psychology, educational sciences, or business administration education, and school to work, or work to school transition episodes. Consequently, we reframed this contextualized self-efficacy in transition-self-efficacy (TSE). In sum, 84 items (16 initial items for self-efficacy, and 7 for SRLC) were included in a pilot questionnaire.

All items were assessed on a five-point Likert scale where (1) indicated strongly disagree and (5) indicated strongly agree. Examples of items in transition self-efficacy (TSE) were “Compared with other students in this master program, I expect to do well” and “I am certain I can understand even the most difficult tasks and topics.” Examples of SRLC were “Even when a learning task is dull, I keep working until I have finished the task” and “I never give up when I am learning something difficult.” Characteristic items in SDLC were “I have clear ideas about what and how I want to learn”, or “If I notice that a certain learning strategy does not work I change my approach.” Examples of items in SDC were “I regularly express my career interests to people who can be of importance for my career”, and “I always keep myself posted about what is important in the eyes of employers.”

Table 2

Components and Subcategories of Self-Directing Learning Capabilities (SDLC) and Self-Directing Career Capabilities (SDC); Raemdonck, 2006, p.77)

Components	**SDLC**	**SDC**

Goal setting	Anticipate on future learning needs	Anticipate changes in the labour market
	Detect knowledge/skill gaps	Detect career opportunities
	Diagnose personal learning needs	Diagnose personal career expectations
	Formulate learning goals	Formulate career goals
		New: Anticipate shared values
Choosing strategy	Collect information about learning opportunities	Collect information about job opportunities
	Select appropriate strategy	Select appropriate strategy
	Develop a learning plan	Develop a career plan
	Identify human and material resources for learning	Identify key persons
Execute strategy	Express learning interests	Express career interests
	Networking to create learning opportunities	Networking and self presentation
	Ask advice to realize learning plan	Ask advice to realize career plan
	Explore learning market and work environment	Explore learning market and work environment
Monitoring and evaluation	Reflect on the self as a learner	Reflect on strengths, preferences, capacities
	Prioritize learning	Prioritize career
	Overcome complexity and negative emotions	Overcome complexity and negative emotions
	Evaluate impact of strategy (result)	Evaluate impact of strategy (result)
	Register progress	Evaluate course of career
	Adjust goals, plan, implementation	Adjust goals, plan, implementation

Procedure

The data were collected in three series. The first data collection was in December 2009. Students in educational science were invited to fill out the questionnaire. This pilot questionnaire was composed of 84 items, and presented online using the BackQuest programme. An invitation of the Dean to participate was sent by email, providing direct access to the web address of the online questionnaire. Two weeks after the first request to fill out the questionnaire, a reminder was sent. The response rate in this first group of respondents was 15% (n = 232). The data of this first group were used to carry out the first factor analyses, confirmatory factor analyses, and to fit constructs in the Rasch model.

A second series of data was collected among advanced psychology students at the Open University in the Netherlands in February 2010, following the same procedure. The response rate in this group was 34% (n = 139). The responses of this second group were combined with the previous responses of the students in educational science to verify whether the same Rasch scales would hold in the combined group. In both groups, no incentives were offered, except for an information reward. Participating students received information about their scores, relative to their fellow students, and a report of the study.

The scales that fit the Rasch rating scale model resulted in a selection of 36 items, which were included in a study resulting in 78 additional responses from master students in business administration of three different business schools in the Netherlands and Flanders.

Analysis

Preliminary analyses.

The data were analyzed in several steps. First, several preliminary analyses were carried out to gain a first insight in the data (i.e. principal axis factor analyses (PAF) and principal component factor analyses (PCA) with Varimax rotation; means, standard deviations, zero order Pearson correlations). Next, we performed more sophisticated analyses, such as confirmatory factor analysis (CFA; Byrne, 2001) and Rasch rating scale modelling in WINSTEPS (Linacre, 2009a) to verify the construct validity. All these preliminary analyses were executed for the final steps: Producing real interval measures, and modelling the constructs using structural equating modelling (SEM) in a causal model, using AMOS version 5 (Arbuckle, 2003).

The first analyses (n = 232) were carried out in SPSS-16. After checking the data for missing values, a PAF with varimax rotation, including all SRLC, SDLC, and SDC items simultaneously was carried out. These items were deliberately combined in the factor analyses to study the Eigen Values (Table 3). Both the PAF and PCA consistently indicated four factors instead of the expected three. The factors were recognized as SDCinformed, SDLC, SDCconscious, and SRLC. The factor solution with four factors explained 46.2% of the variance in the items.

Table 3

Eigen Values and Explained Variance in the Exploratory Factor Analyses (n = 232)

Factors	Eigen value	Explained variance in %	Number of items
F1 SDCinformed	6.52	21.08	10
F2 SDL	3.56	11.47	10

F3 SDCconscious	2.48	8.00	6
F4 SRL	1.80	5.77	5

Transition self-efficacy (TSE) was analyzed separately in a PCA with Varimax rotation, since its factor structure was not subject of discussion. The one factor solution of TSE with 11 items left of the original 16 had an Eigen Value of 4.75, explaining 43.17 of the variance. Item means and standard deviations were calculated (Appendix A[1]). As far as CFA analyses are concerned, fit statistics of the measurement models are provided in Appendix B. Factor loadings are displayed in a separate content validity section (Appendix C) to facilitate comparisons with previous research using CFA. After the analyses in SPSS-16 and CFA, the analysis was continued in the Rasch rating scale model in Winsteps (Linacre, 2009a).

The next section describes fundamental measurement theory, and first explains in practical terms the added value of the Rasch model. The second section concerns fundamental measurement theory, and how to apply the Rasch model for scale diagnosis. Finally, the third section concerns the results of the Rasch analyses that will help us to address our second research question.

Measurement theory: the Rasch model.

Measurement issues in psychological research.

Bond and Fox (2001; 2007) criticize the common practices in the social sciences to use qualitatively ordered raw scores in their summed-up form as if it were interval scales that are required for most statistical analyses. Bond and Fox point out that several serious misapprehensions are made in scale constructions that start with the underlying assumptions made during measurement decisions. According to these authors insufficient care is given to the fundamental measurement question from the operationalization stage to the measurement procedure, namely whether one unit of our units of analysis is the same as the next one. Since the social sciences neglect this issue, the interpretations of their analyses can only be as good (or as dubious) as the quality of their measures. Differences in capabilities of persons should be independent from the items that are scored by the same persons. Bond and Fox (2001) provide several examples, which we will replicate and translate into a real scale in this study, such as self-directed learning capabilities (SDLC). In the procedure of constructing a measurement instrument such as SDLC, the assumption is that high total scores on SDLC indicate a high level of personal agency in managing learning sources, while low total scores on SDLC indicate a low level. Items are scored on a 5-point scale: 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree) and 5 (strongly agree). Let us take five (already Rasch validated) items from SDLC:

Item 27 I am well informed about excellent providers of education programs in my field.

Item 23 I aim to achieve the top level in my profession.

Item 16 If I notice that a certain learning strategy does not work I change my approach.

[1] All data in the appendices are based on the whole sample, $n = 449$

Item 32 I want to perform well, even if I do not like the learning task or theory.

Item 25 I am committed to continuously increase my competencies.

Next, we will use the map of observed raw scores for two different persons A (bold) and B (italic, and grey-marked) on these items (Figure 1).

```
-3    -2     -1      0      1      2      3      4      5      6
|-----+-----+-----+-----+-----+-----+-----+-----+-----|  NUM   ITEM
|                 1  2    3   4         5             |    7  Q27
|                                                     |
|                1  2    3  4       5                 |    5  Q23
|                                                     |
|                                                     |
|                                                     |
|                                                     |
|                                                     |
|            1  2      3  4      5                    |    3  Q16
|                                                     |
|          1        2 3    4     5                     |    8  Q32
|          2      3     4     5                        |    6  Q25
|-----+-----+-----+-----+-----+-----+-----+-----+-----|  NUM   ITEM
-3    -2     -1      0      1      2      3      4      5      6
```

Figure 1. Hypothesized persons A and B in a real Rasch map in empirical item category functions (WINSTEPS, Linacre, 2009a)

For hypothesized person A, with the scores 5-5-4-3-2 traditional practice using Likert-type scales is assigning person A a summed up score 19 on the SDLC-scale, which will be used in subsequent statistical analysis. Yet, person B (grey marked) who responds with 3, 4, 3, 4, and 5, is also assigned the same summed up score of 19, though based on a completely different response pattern. Fox (Bond & Fox, 2001) explains that two assumptions are made, by summing up the ratings: a) each item contributes equally to the SDLC construct and b) each item is measured on the same interval scale. With respect to assumption “a” Figure 1 illustrates that it makes no sense to assume that all items contribute equally to the SDLC construct. The vertical order of the items reflects item difficulties, which are clearly distinct. It is easier to agree with item 25 at the bottom of the item difficulty hierarchy than with item 23 or 27 in the top. Strong agreement with item 27 (I am well informed about excellent providers of education programs in my field) indicates more SDLC than strong agreement with item 25, since item 25 can easily be endorsed by most respondents. However, strongly endorsing item 27 indicates that the respondent does not consider being well informed about excellent education providers as difficult, which is uncommon. Item 23 is somewhat less difficult than item 27, but still substantially more difficult than the items 16, 32 and 25. The item is strongly indicative for ambitious goal setting. The order of the items from bottom to top represents different difficulty levels in endorsing the SDLC items, although the items are equally discriminative. Bond and Fox (2001) explain that in such situations data have to be analyzed in such a way that the person scores on the items reflect person capabilities on items that indirectly measure the continuum of an underlying psychological phenomenon (unidimensionality).

The second assumption, underlying summing up the rates is based on considering the scale as an interval scale, presuming that the distance between each scale distance (from 1 to 2; from 2 to 3; from 3 to 4, and from 4 to 5) is uniform within and across items. There can be quite large psychological differences from endorsing disagreement to agreement and this lack of linearity within items can also be manifested across items. The value of the distances between the rating scale categories may differ for each item. Agree and strongly agree may be close for some items but not for others, which is illustrated by the real observed distances in the SDLC-scale (see Figure 1). The map reveals that the value of distances between the rating scale categories differs for each item and also from item to item within a scale. It emphasizes that a Likert-type scale is not more than a qualitatively ordered (ordinal) scale, where scores have to be mathematically transformed before they can be used as interval measures. It also demonstrates that when Likert scales are summed up (which is not allowed for ordinal scales) the summed up score will be distorted and biased. It will negatively affect all subsequent statistical analyses that assume that the scale will have interval properties. The bias is the consequence of the ungrounded assignment of summed up scores to values that are not reflecting uniform distances, and of scores on items that contribute unequally to the latent factor. In such instances only the Rasch model provides a mathematically sound alternative (Bond & Fox, 2001, 2007). The Rasch model defines the measures, based on the properties of the underlying psychological variable, instead of data that are defining the measurement model.

Rasch: theory and practice.

To determine the validity of the measures Rasch modelling techniques are used. Rasch analysis provides for two sets of guidelines (Bond & Fox, 2001, p. XX). First, the researcher must assess whether all the items work together to measure a single latent variable. Although Likert-type items have not been constructed originally in a hierarchy to put items and persons on a single scale ranging from very easy to very difficult (or from low abilities to high abilities), studies have shown that the Rasch model holds and is robust, even if such assumptions are not made during test item design (Forsyth, Saisangjan, & Gilmer, 1981).

Construct validity is the approximate truth of the conclusion that an operationalization accurately reflects an underlying psychological construct (Trochim, 2002). Claiming construct validity essentially is claiming that the observed pattern corresponds with the theoretical pattern. Trochim (2002) calls this process pattern matching and considers it as the heart of construct validity. It relates to the question whether the operationalized construct is one-dimensional. From the perspective of Rasch modelling construct validity is defined as the relation between the difficulty order of the items produced by the manner in which persons respond to the items and their content (Bond & Fox, 2001). Rasch modelling informs the researcher about the relative value of every one of the person capability (or endorsability) measures, and item difficulty estimates (Näring, Hoogduin, & Keijser, 2004).

Second, in Rasch analysis the validity of a scale is assessed through examination of the fit of the model (Bond & Fox, 2001, 2007). Fit statistics are derived from a comparison of the expected patterns and the observed patterns of response. These fit statistics are used as an assessment of the validity of the model-data fit, and as a diagnosis for individual item or person abilities. Each scale in this study has been analyzed for its fit in the Rasch model. Stringent Rasch criteria have been applied to all items that are combined in a scale to measure a particular underlying variable.

All Rasch analyses have been executed in WINSTEPS that uses a joint maximum likelihood estimation (JMLE; Linacre, 2009a). The transformation of ordered qualitative observations into additive measures is a Rasch model. Rasch models are logit-linear models or log-linear models. The polytomous[2] Rasch rating scale model that is applied in the current study uses the following additive transformation (Linacre, 2009b, p. 32-33):

$$\log(Pnij/ Pni(j-1)) = Bn - Di - Fj$$

where Pnij is the probability that person n encountering item i is observed in category j; Bn is the "ability" measure of person n; Di is the "difficulty" measure of item i, the point where the highest and lowest categories of the item are equally probable. Fj is the "calibration" measure of category j relative to category j-1, the point where categories j-1 and j are equally probable relative to the measure of the item. No constraints are placed on the possible values of Fj.

The model fit is evaluated using mean square infit and outfit statistics (Linacre, 2009b). The mean square infit statistic is a t-standardized information weighted mean square statistic, insensitive to unexpected responses to items far from the person’s level of ability, but sensitive to unexpected responses to items near the person’s measure level. The outfit statistic is a t-standardized outlier-sensitive mean square fit statistic, which is more sensitive to unexpected behaviour by persons on items far from the person’s ability level. The expected value of these statistics is one. Values, substantially less than one, indicate dependence in the data. Values, substantially greater than one indicate noise (Linacre, 2009b).

A first diagnosis of the scale is based on the average item (Table 4; A) and average person (B) infit mean square and outfit mean square, supported by the average item measure (A_a), which is fixed at zero, and the average person measure (B_a), which may be negative, positive or approximately be zero. A negative average person measure indicates a difficult test for the group of persons under investigation; a positive average person measure reflects a relatively easy test for the sample of persons. In combination with the standard deviations (A_d) for the average item measure (A_a) and the person average measure (B_a) and its standard deviation (B_d), and the respective in- and outfit mean squares (A_b, A_c, and B_b, B_c) and standard deviations (A_e and A_f, and B_e and B_f) the Rasch analysis provides a first impression about the measurement quality of the scale.

[2] Polytomous means a scale with more than two categories.

Table 4

Guideline for Reported Rasch Fit Statistics

Item	**H** Infit	**I** Outfit	**J** Measure	**K** Error	**L** PTMEA	Miscellaneous
nr	.50 < infit < 1.50	50 < infit < 1.50	Most difficult			
nr						
nr						
nr						
nr						
nr			Least difficult			
All items (**A**)						
Mean	A_b	A_c	A_a			Person Reliability (**C**)
SD.	A_e	A_f	A_d			Person Separation (**D**)
All persons (**B**)						Item Reliability (**E**)
Mean	B_b	B_c	B_a			Item Separation (**F**)
SD	B_e	B_f	B_d			Cronbach alpha (**G**)

The Rasch model provides two reliability measures and separation measures: a person reliability (C) and person separation measure (D), and an item reliability (E) and item separation (F) measure. The person and item reliabilities are based on the ratio of true (person or item) variance divided by the observed (person or item) variance (Linacre, 2009b). The Rasch person reliability functions as a lower bound for Cronbach alpha (G; Linacre, 2009b). The item reliability that accompanies the person reliability, offers the opportunity to evaluate the reliability of item difficulty separately from the person reliability, in contrast to Cronbach alpha that only evaluates person reliability (Fox & Jones, 1998). With respect to item reliability, values >.90 clearly indicate a one-dimensional variable, measured by internally consistent items.

The separation indices are indicators of the scales' person and item discrimination functions, in other words: how well the scale discriminates between persons with high (endors)abilities or low (endors)abilities, and how well does the scale discriminates between easy and difficult items. The separation index is the ratio of the unbiased estimate of the sample standard deviation to the root mean square measurement error of the sample. A computation is used to transform the person and item separations into person ability strata and item difficulty strata that are distinguished by the scale. The calculation is: four times the separation index plus one, divided by three [(4G+1)/3], indicating regions of the scale, whose centres are separated by logit distances greater than can be explained by measurement error (i.e. beyond three standard errors). A scale must reach out to at least two item difficulty strata or person ability strata to meet the minimum requirement of two distinct difficulty strata that must be distinguished by the persons and to meet the same minimum criterion of two distinct person ability strata that must be distinguished by the items (Curda, 1997). These calculated strata will be reported after modelling each Rasch scale.

Even more important, the Rasch model provides in- and outfit means squares for each single item (Table 4, H and I) included in the model, supporting the evaluation of each item within the construct. This is particularly relevant during instrument development. The scale can also be investigated focusing on properties such as a well balanced spread of items along the continuum of the latent variable (see the Figures 2-6). Each item is equipped with an item estimate (logit or measure; J), and an individual item error (K), which makes it possible to judge the reliability of each item's difficulty in replications in other samples. Rasch item calibrations are invariant within standard error estimates (Fox & Jones, 1998). This feature makes it possible to use calibrated item measures as anchoring measures for comparisons of tests, scales or groups (Beltyukova, Stone, & Fox, 2004; Linacre, 2009b; Bode, 2001; Smith & Dupeyrat, 2001; Taherbhai & Young, 2004; Wolfe, 2000; Yu & Osborn Popp, 2005). The same principles can be applied to calibrated person measures. PTMEA (L), the point-measure correlation, reflects the correlation of an item's (or person's) measures with the measures of the encountered persons (or items; Linacre, 2009b).

The item measures (J) in a Rasch rating scale model have to be in ascending order, from easy to difficult, just like the step calibration measures (thresholds[3]). The latter (not reported in the Tables, but diagnosed) reflect the probability of each step in the rating scale in the model (Linacre, 2009b). It is expected that the relative difficulties of the categories in each item should be relatively similar across items. The probability that an individual gives a particular response to a question is estimated in two terms. The person parameter estimates the location of the person on the underlying construct for which the question has been designed to assess. Rating scale step measures and individual item difficulty estimates are used to represent the transition within that question between one response and the next (e.g. the step from agree to totally agree or the step from disagree to totally disagree). Rasch computes the probability of a response as a function of a person's estimated (endors)ability, and the estimated difficulty related to the steps between adjacent response alternatives, associated with a question. Response category probability curves illustrate the points along the ability scale, indicated by the intersections of the curves, where the likelihood of a particular response changes to the next step (Curda, 1997). The more a person's ability exceeds the item difficulty, the greater the difference and the higher the person's probability of success (Wright & Mok, 2004). When item difficulty exceeds person's latent ability, resulting in a negative difference, than the person's probability of success is less than .5 (Wright & Mok, 2004). A Rasch model assumes that all items are equally effective in discriminating among respondents. If the response patterns of items and persons fit the requirements of Rating Scale Model, then the construct has been well-defined and measured and test- and person-

[3] To gain more insight in the thresholds, appendix D provides illustrations of Rasch-Andrich thresholds and Rasch-Thurstone thresholds, and how they are used in scale diagnoses. The Rasch-Andrich threshold is part of the Rasch Rating Scale formula (Fj).

independent measures are obtained (Linacre, 2009b; Wright & Mok, 2004). Adding items will not change a person's capability (or endorsability) position on the continuum, but only refine and improve the test quality (Wright & Stone, 1999). The Rasch Rating Scale Model calculates all parameters necessary for construct validation and can transform a fitting qualitatively ordered Likert-type scale into a desired interval scale (Wright & Mok, 2004), providing the means (measures) for unbiased statistical analyses (Bond & Fox, 2001). Via mathematical computations measures on an interval scale are produced, expressed in the natural log odds of both persons and items (logits) along the same scale (Bond & Fox, 2001; Linacre, 2009b). Logits typically range from - 4 to + 4, with increasing person estimates indicating higher ability and increasing item estimates indicating greater difficulty. As described and (partly) illustrated in Table 4, the Rasch model provides a richness of indices to evaluate the measurement quality of item content, person responses, scale construct validity, and reliability of the estimates.

Results in the Rasch rating scale model.

Each construct has been scrutinized in the Rasch rating scale model. While the factor analyses identified one self-regulating construct and three self-directing constructs, the Rasch model did the same, however, in remarkably different scale compositions. This very issue indicates that in common non-Rasch measurement methods, person characteristics and item qualities can be strongly intertwined. In other words, specifically during instrument development, in non-Rasch methods the designer cannot identify whether person capabilities or features of the instrument are measured. Below we present the Rasch results for each identified construct, based on the data of the complete sample (*n* = 449), starting with the two differential SDC constructs, and the SDLC construct. The self-directing constructs will be followed by the Rasch results of self-regulating SRLC and TSE.

Table 5

Self-Directed Career Capabilities_Informed (SDCinformed; *n* = 449)

Item	Infit	Outfit	Measure	Error	PTMEA	Miscellaneous	
45	.98	.97	.53	.07	.74		
58	1.14	1.17	.20	.08	.67		
53	.90	.90	.15	.07	.76		
54	.93	.93	-.13	.07	.76		
43	.99	.97	-.21	.07	.74		
60	1.04	.99	-.53	.08	.70		
All items							
Mean	1.00	.99	.00	.08		Person Reliability	.80
SD.	.08	.09	.34	.00		Person Separation	2.03
All persons						Item Reliability	.95

Mean	.99	.99	.35	.74	Item Separation	4.33
SD	.83	.85	1.72	.18	Cronbach alpha	.84

Characteristic for the SDCinformed scale is that the scale predominantly expresses cognitive and behavioural activities that are needed to gather information about the labour market and job opportunities.

The data fit the model well with global mean square infit and outfit statistics for items of 1.00 and .99 respectively, and for persons of .99 and .99 respectively. The average measures and step calibration measures are both in ascending order. The average measures are the average ability across all items of persons responding in that response category. Ordered step calibration measures indicate that each response category is the most probable response at some point at the response continuum. The scale has an item separation of 4.33 and an item reliability .95, and a person separation of 2.03 and a person reliability of .80. This implicates that the scale identifies 6.11 distinct item difficulty strata that the participants distinguish and 3.04 distinct SDC-informed person capability strata, distinguished by the items. Both are well above the criteria of two strata needed for the scale to be useful in distinguishing individuals with high and low SDCinformed abilities. This scale can even discriminate three different person capability groups (low-moderate and high) in SDCinformed capabilities. The range of item calibrations is from -0.53 *logit* (item 60) to 0.53 *logit* (item 45). The ordering of the items provides evidence of the construct validity, given that the items seem to be logically ordered from least to most difficult. The person and items map of the variable is depicted in Figure 2.

The average item measure is 0.00 *logit* (*SD* = 0.34) and the average person measure is 0.35 *logit* (*SD* = 1.72). The most difficult item measure (item 45, 0.74 *logit*) exceeds the average person measure, indicating that the scale adequately assesses the higher end of the continuum in SDCinformed. The same goes for the middle and lower part of the continuum.

The variance that is explained by the measures is 53.8%. 34.4% is explained by the persons, and 19.4% by the items.

The easiest items to endorse are items that reflect expectancies (clear subjective ideas) about the future career, and items on being well informed about career opportunities. Other items that reflect the gathering of information are relatively easy. Identifying persons that can be approached to increase the chance of obtaining a desired job (item 45) seems to be the most difficult issue for the respondents. The latter might indicate that professionals can be supported in their career development when they learn to identify and recognize the roles fulfilled by different persons in different communities.

```
--------------------------------------------------------------------------------

           PERSON - MAP - ITEM
               <more>|<rare>
    7              .  +
                      |
                      |
                      |
                      |
    6              #  +
                      |
                      |
                      |
                      |
    5              .  +
                      |
                      |
                  .#  |
                      |
    4                 +
                 .## T|
                      |
                      |
                      |
    3            ###  +
                      |
                      |
          ##########  |
                      |
    2                S+
             .######  |
                      |
                      |
         .##########  |
    1                 +
            ########  |
                      |T Q45
        .########### M|S
                      |  Q53    Q58
    0     ##########  +M
           .########  |  Q43    Q54
                      |S
            .#######  |T Q60
                      |
   -1       .#######  +
            .#######  |
                     S|
               .####  |
                      |
   -2         .#####  +
                      |
                 .##  |
                      |
                      |
   -3              . T+
                      |
                   .  |
                      |
                      |
   -4              .  +
                      |
                      |
                   .  |
                      |
   -5              .  +
               <less>|<frequ>
 EACH "#" IS 4. EACH "." IS 1 TO 3
```

Figure 2. Person and item map SDCinformed

Table 6

Self-Directed Career Capabilities_Conscious (SDCconscious; *n* = 449)

Item	Infit	Outfit	Measure	Error	PTMEA	Miscellaneous	
56	.94	.95	1.14	.07	.69		
63	1.12	1.13	.85	.07	.62		
59	.79	.77	.35	.07	.73		
47	.80	.78	.23	.06	.73		
61	1.17	1.18	-.14	.07	.56		
46	1.09	1.08	-.19	.07	.63		
55	1.08	1.05	-.58	.07	.60		
52	.94	.89	-.81	.07	.67		
66	1.10	1.11	-.85	.09	.52		
All items							
Mean	1.00	.99	.00	.07		Person Reliability	.82
SD.	.13	.14	.66	.01		Person Separation	2.12
All persons						Item Reliability	.99
Mean	.99	.99	.66	.56		Item Separation	8.93
SD	.68	.70	1.35	.13		Cronbach alpha	.84

SDCconscious is a SDC-scale that can be distinguished from the SDCinformed scale by its more internalized focus on the career. In contrast with SDCinformed, SDCconscious reflects deeper reflection on career development, and includes the recognition of social capital in achieving personal career goals.

The data fit the model well, with global mean square infit and outfit statistics for items of 1.00 and .99 respectively, and for persons of .99 and .99. The average measures, and step calibration measures for all items are in ascending order. The scale has an item separation of 8.93, and an item reliability .99. The person separation is 2.12 and the person reliability .82. This means that the scale identifies 12.24 distinct item difficulty strata that the participants distinguish and 3.16 distinct SDCconscious strata, distinguished by the items. Both are well above the criteria of two strata needed for the scale to be useful in distinguishing individuals with high or low SDCconscious. The scale can even distinguish high, medium, and low SDCconscious-abilities. The range of item calibrations is from -0.85 *logit* (item 66) to 1.14 l*ogit* (item 56), which provides evidence of the construct validity. The items seem to be logically ordered from least to most difficult. The person and items map is depicted in Figure 3.

The average item measure is .00 *logit* (*SD* =.66) and the average person measure is .66 (*SD* = 1.35). The most difficult item measure (item 56, 1.14 *logit*) and the next most difficult item (item 63, 0.85 logit) both exceed the average person measure, indicating that the scale also assesses the higher end of the SDCconscious continuum.

The variance that is explained by the measures is 49.5%; 27.8% is explained by the persons and 21.7% by the items.

The easiest items to endorse (item 66, item 52, and item 55) are about checking the sense of reality of personal career goals, the usefulness of having many contacts, and personal career plans. Pursuing ambitious career goals, and consulting others on how to improve personal effectiveness are also items that are perceived as relatively easy to endorse. Being ambitious, consulting others, informing the network about one's career plans and using the network to improve the chances of obtaining a desired job are perceived as moderately difficult. However, the most difficult items concern negotiation with the (future) supervisor, and keeping the network posted on the career plans.

```
--------------------------------------------------------------------------------
           PERSON - MAP - ITEM
               <more>|<rare>
    6              .  +
                      |
                      |
                      |
                   .  |
    5                 +
                      |
                   #  |
                      |
                      |
    4              .  +
                      |
                  .#  |
                     T|
                  ##  |
    3                 +
                 .##  |
                      |
                .###  |
                      |
    2         ###### S+
              ######  |
                      |
        .##########   |T
          #########   |  Q56
    1                 +
           .#######   |  Q63
          .######## M|S
         .#########   |  Q59
          #########   |  Q47
    0      .#######   +M
             .#####   |  Q46    Q61
              .####   |
                .## S|S Q55
                 ##   |  Q52    Q66
   -1         .####   +
                 ##   |
                 .#   |T
                 .#   |
                  .   |
   -2             .  T+
                  .   |
                      |
                  .   |
                  .   |
   -3                 +
                  .   |
                      |
                      |
                      |
   -4                 +
                  .   |
                      |
                      |
                      |
   -5                 +
                      |
                      |
                      |
                  .   |
   -6                 +
               <less>|<frequ>
 EACH "#" IS 4. EACH "." IS 1 TO 3
```

Figure 3. Person and item map SDCconscious

Table 7

Self-Directing Learning Capabilities (SDLC; *n* = 449)

Item	Infit	Outfit	Measure	Error	PTMEA	Miscellaneous	
27	.99	1.00	.81	.06	.57		
9	1.06	1.10	.59	. .06	.54		
23	.90	.91	.55	.06	.64		
16	1.01	.98	-.19	.08	.45		
18	1.00	1.00	-.35	.07	.50		
12	.99	.97	-.38	07	.53		
25	.87	.82	-.51	.09	.57		
32	1.16	1.17	-.51	07	.40		
All items							
Mean	1.00	.99	.00	.07		Person Reliability	.62
SD.	.09	.10	.52	.01		Person Separation	1.28
All persons						Item Reliability	.98
Mean	1.00	.99	1.02	.58		Item Separation	6.95
SD	.74	.75	.97	.13		Cronbach alpha	.65

.

The scale is predominantly composed of newly formulated items. The fit of the model is adequate, with global mean square infit and outfit statistics for items of 1.00 and .99 respectively, and for persons of 1.00 and .99. The range of the SDLC scale is from -0.51 *logit* to 0.81 *logit*. The top end of the SDL scale (item 27, 0.81 *logit*) is 0.21 *logit* removed from the average person measure, which is 1.02 *logit* (*SD* = .97). It means that persons too easily endorse the current SDLC scale as a whole, and that the top end of the continuum in SDLC is not yet accurately measured.

The person reliability is .62; its separation however is still sufficient, 1.28. The item reliability .98 is acceptable, just like the item separation of 6.95. The SDLC scale identifies 9.60 distinct item difficulty strata that the participants distinguish and 2.04 distinct SDLC-person strata, distinguished by the items. The item strata are both above the required two, and the person capabilities are just sufficient to distinguish less and more capable persons in SDLC. A weakness in the construct is the item difficulty 'jump' between item 16 and item 23 of 0.74 logit.

However, during all previous factor analyses, CFA's, and series of Rasch analyses SDLC appeared to be the most problematic construct to compose, since all available items together were not able to measure a substantial part of the underlying continuum (while simultaneously, demonstrating a Cronbach alpha of .83 in the original 35 items edition). Originally 21 items were purposefully and theory-based designed for SDLC to complete the already existing 14 items that were designed by

Raemdonck (2006), resulting in a SDLC item pool of 35 items. However, more than 51% of these 35 items showed Rasch-Andrich threshold problems, and another 14 % showed misfits. The complete pool of items was not able to measure person responses beyond a range of -1 and +2 logit (instead of the common -4 to +4 logits). Whereas the category range was very small, the difficulty hierarchy was reduced because the two originally most difficult SDL items fell outside the acceptable range of in- and outfit mean square fit statistics for items. Remaining items that did fit the Rasch model were relatively easy to endorse. Obviously, our definitely composed SDLC-scale shows such shortcomings, though threshold problems and misfits do not appear in this definite version. The variance that is explained by the measures is 37.6%; 17.9% is explained by the persons, and 19.8% by the items. The person and item map of SDLC is presented in Figure 4.

The easiest items to endorse are items concerning the intention to perform well (even in unattractive tasks), and the willingness to develop one's competences. Items that refer to what has to be learnt, on getting involved in projects that offer many learning opportunities, and the need to change one's learning strategy are also easy to endorse. However, items that represent the ambition to become a top professional, or being well-informed about professionalization courses and good training institutes are moderately difficult to endorse.

```
-------------------------------------------------------------------------------

           PERSON - MAP - ITEM
               <more>|<rare>
    6             .  +
                     |
                     |
                     |
                     |
                     |
    5             .  +
                     |
                     |
                     |
                     |
                  .  |
    4                +
                     |
                  .  |
                     |
                     |
                 .#  |
    3               T+
                .##  |
                     |
                     |
               .###  |
                     |
    2        .##### S+
                     |
          #########  |
                     |
        .##########  |
                     |
    1 .############ M+T
      .############  |  Q27
                     |  Q9
           .#######  |S Q23
            #######  |
             .#####  |
    0               S+M
               .###  |  Q16
               .###  |  Q12    Q18
                  .  |S Q25    Q32
                 .#  |
                  .  |
   -1               T+T
                  .  |
                  .  |
                     |
                     |
                     |
   -2             .  +
                     |
                     |
                  .  |
                     |
                     |
   -3                +
               <less>|<frequ>
 EACH "#" IS 5. EACH "." IS 1 TO 4
```

Figure 4. Person and item map SDLC

Table 8

Self-Regulating Learning Capabilities (SRLC; n = 449)

Item	Infit	Outfit	Measure	Error	PTMEA	Miscellaneous	
38	.97	.95	.97	.09	.70		
22r	1.18	1.23	.61	.08	.63		
26	1.02	.99	-.02	.08	.67		
28	.83	.80	-.75	.09	.72		
15r	.94	.93	-.80	.08	.69		
All items							
Mean	.99	.98	.00	.08		Person Reliability	.70
SD.	.11	.14	.71	.00		Person Separation	1.52
All persons						Item Reliability	.99
Mean	.98	.98	1.43	.91		Item Separation	8.34
SD	.85	.95	1.27	.25		Cronbach alpha	.74

The fit of the model is adequate with global mean square infit and outfit statistics for items of .99 and .98 respectively, and for persons of .98 and .98. The global person infit- and outfit mean squares show a slight overfit. The average measures and step calibration measures are both in ascending order, and the ordered step calibration measures in all included items indicate that each response category is the most probable response at some point at the response continuum.

The scale has an item separation of 8.34, and an item reliability .99. The person separation of 1.52 is adequate for such a small scale, just like the person reliability of .70. It implicates that the scale identifies 11.45 distinct item difficulty strata that the participants distinguish and 2.36 distinct SRLC-capability strata, distinguished by the items. Both are above the criteria of two strata needed for the scale to be useful in distinguishing individuals with high and low abilities in SRLC. The range of item calibrations is from -0.80 (item 15r) to 0.97 *logit* (item 38) l*ogit*. The ordering of the items provides evidence of the construct validity, given that the items seem to be logically ordered from least to most difficult. The person and items map of SRLC is depicted in Figure 5. The variance that is explained by the measures is 53%; 33% is explained by the persons, and 20% by the items.

The average item measure is .00 l*ogit* (*SD* .71) and the average person measure is 1.43 *logit* (*SD* 1.27). The most difficult item measure (item 38, 0.97 l*ogit*) is well below (0.46 *logit*) the average person measure, indicating that the scale does not yet measure the higher end of the continuum in SRLC. The current SRLC scale is too easy to endorse by professionals. For future scale development it can be recommended to empower the perseverance items to overcome complexities, and negative emotions, to more equivocally measure the higher end of the continuum.

The easiest items to endorse are items that refer to remaining focused and persistent, and to accomplishing education programs. Not getting distracted and being persistent in more complex or difficult tasks seem to be moderately difficult issues for professionals.

```
--------------------------------------------------------------------------------

          PERSON - MAP - ITEM
              <more>|<rare>
   6            .#  +
                    |
                    |
               ###  |
                    |
   5                +
                   T|
                    |
                    |
             .####  |
   4                +
                    |
                    |
                    |
         .######## S|
   3                +
                    |
                    |
        .#########  |
                    |
   2                +
                    |
      .###########  |
                   M|T
                    |
   1   .##########  +  Q38
                    |S
                    |  Q22R
        .#########  |
                    |
   0         .####  +M Q26
                   S|
            .#####  |
                    |
               .##  |S Q15R   Q28
  -1                +
                .#  |
                    |T
                .#  |
                    |
  -2            .# T+
                    |
                 .  |
                    |
                    |
  -3             .  +
              <less>|<frequ>
EACH "#" IS 6. EACH "." IS 1 TO 5
```

Figure 5. Person and item map SRLC

Table 9

Transition Self-Efficacy (TSE; *n* = 449)

Item	Infit	Outfit	Measure	Error	PTMEA	Miscellaneous	
74r	.92	.95	.98	.07	.63		
72	.85	.86	.56	.07	.65		
69	1.10	1.10	.30	.08	.44		
84	.83	.81	-.01	.08	.65		
82	.96	.96	-.08	.07	.61		
76r	1.00	1.00	-.18	.06	.61		
80	1.29	1.29	-.74	.07	.38		
79r	1.01	1.05	-.83	.07	.55		
All items							
Mean	1.00	1.00	.00	.07		Person Reliability	.67
SD.	.14	.14	.58	.01		Person Separation	1.43
All persons						Item Reliability	.98
Mean	1.00	1.00	.75	.59		Item Separation	7.84
SD	.76	.77	1.05	.13		Cronbach alpha	.71

As described in the theoretical section, self-efficacy is the most contextualized self-concept. To tune the concept with our context of adult students, who either are in the transition from school to work, or from work to school, we modified several items in the original Pintrich et al. (1991, 1993) scale and formulated specific professional competence items for the professions. Consequently, we relabelled the construct into transition self-efficacy, which is supposed to be an accurate self-efficacy measure for boundaryless careers in all professions.

The global mean infit and outfit statistics for items and persons of 1.00 and 1.00 indicate a good fit. The measures are in ascending order and the step calibration measures are well structured in all items. Both provide evidence for the construct validity of the scale.

The range of the transition self-efficacy scale is from -0.83 *logit* (item 79r) to 0.98 *logit* (item 74r). The most difficult item (item 74r; 0.98 *logit*) exceeds the average person measure (0.75 l*ogit*; *SD* = 1.05), indicating that the scale also assesses the higher end of the TSE-continuum, as well as the medium and lower end. The TSE scale identifies 10.79 distinct item difficulty strata that the participants distinguish and 2.24 distinct TSE-capability strata, distinguished by the items. Both are well above the criteria of two strata needed for the scale to be useful in distinguishing individuals with high and low TSE abilities. The variance that is explained by the measures is 39%; 18% is explained by the persons, and 21% by the items.

The easiest items to endorse are items that represent worrying about one's key professional competences (reversed), endorsing that one is successful in one's (future) profession, worrying about study results (reversed), and a statement that one is able to understand difficult tasks. Moderately difficult are items that either refer to social comparisons or to being able to use theory in practice. The most difficult issue in this scale is a social comparison item that states that one does not outperform other professionals (reversed). The latter might be a cultural issue, since in the Netherlands it is not socially accepted to publicly acknowledge that one outperforms others.

```
--------------------------------------------------------------------------------

           PERSON - MAP - ITEM
               <more>|<rare>
    5               .  +
                       |
                       |
                       |
                    .  |
                       |
    4                  +
                    .  |
                       |
                       |
                    #  |
                       |
    3                  +
                  ## T|
                       |
               .####   |
                       |
                  .#   |
    2                  +
            .####### S|
                       |
      .###########     |
                       |
       .##########    |T
    1   ###########    +  Q74R
                     M|
     #############     |
                       |S Q72
     .############     |  Q69A
       .##########     |
    0                  +M Q82    Q84
            .#######   |  Q76R
               .#### S|
               #####  |S
                       |  Q80
                 .##   |  Q79R
   -1            .##   +
                    #  |T
                     T|
                    .  |
                    #  |
                    .  |
   -2                  +
                    .  |
                       |
                       |
                    .  |
                       |
   -3                  +
               <less>|<frequ>
 EACH "#" IS 4. EACH "." IS 1 TO 3
```

Figure 6. Person and item map TSE

Conclusion/discussion Rasch analysis.

In the preliminary analyses we validated the differential self-regulating and self-directing constructs in the Rasch model to address our second research question "Do the self-regulating and self-directing constructs fit the Rasch rating scale model, to produce Rasch interval measures?" We can confirm this question for all identified five constructs. All constructs are able to distinguish at least two person capability strata (both SDC constructs even three) and more than two item difficulty strata.

All analyses have clearly revealed five different constructs (instead of the expected four). One of the merits of these analyses is that self-directing career capabilities is identified as a construct that is composed of two different dimensions. Both career constructs were consistently depicted as two separate factors in the original educational science sample (n = 232), and in the whole sample (n = 449), suggesting that the two dimensions are inclined to vary differently. Both constructs provide deeper insights in mechanisms that drive career opportunities. The SDCinformed construct highlights that being well-informed on labour market issues is an important competence to cope for oneself in an increasingly competitive labour market. In addition, the SDCinformed competence appears to be highly trainable, although current career training and workshops tend to neglect labour market developments, and focus predominantly on personal issues. SDCconscious has the characteristics of a personalized construct. The two different dimensions inform us on differential difficult issues in career trajectories, even for experienced professionals. SDCinformed reveals that identifying persons that can be approached to increase the chance of obtaining a desired job is perceived as a very difficult issue, which suggests that professionals can be supported in their career development when they learn to identify and recognize the roles fulfilled by inbounds, insiders and boundaries in different communities (Lave & Wenger, 1991). SDCconscious reveals that professionals are uncomfortable regarding negotiating skills, and might be made aware that informing peers or supervisors about career plans can be a natural part of informal or task-oriented conversation. A more negotiating style of communication doesn't seem to be well-integrated in the competence repertoire.

Some issues regarding future scale improvement should be considered:

a. The SDLC construct appeared to be a problematic construct, since many SDLC items are characterized by extremely small category distances that very easily get distorted, and together, are unable to construct an accurate scale. In this study we deliberately only selected items without misfits and disordered Rasch-Andrich thresholds, in order to position a person's endorsability as accurate as possible. We can however consider operationalizing a subgroup of items as dichotomous items (yes/no), and investigating whether such a subgroup of items can be integrated in a scale that also contains polytomous items. It might be that a combination of dichotomous and polytomous items is a more realistic representation of SDLC. The gap, in the middle of the current scale, might be overcome by such a procedure.

b. SRLC can still be improved to measure the top end of the continuum. Since specifically overcoming complexities, negative emotions, and the adjustment of plans are distinguishing (and more difficult) issues in this scale, additional items for these components can be added to the scale in future.

Three scales (SDCinformed, SDCconscious, and TSE) already represent the higher, middle, and lower part of the continuum. Although SDLC and SRLC can distinguish two person capability strata, when these particular scales are interpreted we have to keep in mind that both constructs do not yet measure the higher end of the continuum.

The Rasch model provides accurate diagnoses about which specific issues are 'difficult' for respondents, informing designers on which training interventions might be needed. In addition, items in scales of differently operationalized, but similar constructs can be anchored on the currently available Rasch measures, allowing measurement comparisons across different empirical studies.

Since all five scales are now validated in the Rasch model, the Rasch interval person measures that are produced will be modelled in the next section, using structural equation modelling. In the next section we will first describe the procedures concerning structural equation modelling.

Modelling the constructs in SEM.

In confirmatory factor analysis (CFA) and in structural equation modelling (SEM) the maximum likelihood method is used to determine the goodness of fit between the hypothesized model and the sample data (Byrne, 2001). SEM produces several goodness of fit statistics, which all provide information about the model's fit. When Σ is the population covariance matrix, S is the sample covariance matrix and θ is a vector that represents the model parameters, then $\Sigma(\theta)$ is the representation of the restricted covariance matrix implied by the model in the population. In SEM the null hypothesis (H_0) is tested that the postulated model holds in the population, $\Sigma = \Sigma\ (\theta)$ and that is why the researcher's intention is *NOT* to reject the null hypothesis (Byrne, 2001). The null hypothesis postulates that specification of the factor loadings, factor variances/covariances and error variances for the model under study are valid. The primary focus of the estimation process in SEM is to yield parameter values such that the discrepancy between the sample covariance matrix S and the population covariance matrix implied by the model [$\Sigma\ (\theta)$] is minimal.

The chi-square statistic (χ^2) represents the likelihood ratio test statistic or the discrepancy between the unrestricted sample covariance matrix S and the restricted covariance matrix [$\Sigma\ (\theta)$]. With respect to the *p*-value of this statistic, the higher the probability, associated with χ^2 the closer is the fit between the hypothesized model and the perfect fit. Byrne (2001) points out that the chi square statistic (χ^2) is very sensitive to sample size and is built on the assumption that the model fits perfectly in the population, which will not be feasible using real world data. Therefore researchers have developed goodness of fit indices that take a more pragmatic approach to the evaluation process (Byrne, 2001). These criteria are used as adjuncts to the χ^2 statistic. Houkes (2002) suggests using goodness of fit criteria that assess the model fit, model comparison and model parsimony.

Adjunctive model fit criteria used in our studies are the Goodness of Fit Index (GFI) and the Adjusted Goodness of Fit Index (AGFI). The GFI is a measure of the relative amount of variance and covariance in the sample covariance matrix (S) that is jointly explained by Σ (the population covariance matrix). The AGFI is an overall measure of fit, which, in contrast with GFI, takes the degrees of freedom into account. Both indeces compare the model with no model at all and range from zero to one. Values close to one are indicative of good fit (Byrne, 2001), but in practice values within the range .90 – 1.0 are considered as indicative for a good fit. AGFI estimates the extent to which the sample variances and covariances fit the hypothesized model, while taking parsimony into account (Houkes, 2002). When AGFI is .85 or higher, the model fit is acceptable, and values around .95 indicate a good fit. Just like the chi-square statistic, both GFI and AGFI are sensitive to sample size.

Comparative indices of fit used in our study are the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI, sometimes also labelled as NNFI). These indices compare the hypothesized model with the null model (assuming zero relationships between variables), as the baseline model. CFI takes sample size into account and measures the improvement in non-centrality in going from the least restrictive to the saturated model. CFI should be higher than .90 (Schumacker & Lomax, in Houkes, 2002). The TLI allows the comparison of models, regarding whether the models are nested in an ordered sequence or not. The index is not sensitive to sample size and, just like CFI, and should be higher than .90. Byrne (2001) however advises to use the cut-off score .95 in both indices.

The fit criterion to assess model parsimony is the Root Mean Square Error of Approximation (RMSEA), which recently has been recognized as one of the most informative criteria in covariance structure modelling (Byrne, 2001). The RMSEA takes into account the error of approximation in the population and considers how well the model, with unknown but optimally chosen parameter values, would fit the population covariance matrix if it were available. The discrepancy, as measured by the RMSEA, is expressed per degree of freedom, which makes the index sensitive to the number of estimated parameters in the model (the complexity of the model). Values less than .05 indicate good fit; values ranging from .06 - .10 represent mediocre fit; values > .10 indicate a poor fit. It is strongly advised to use confidence intervals in practice. For example, when a researcher is confronted with a small RMSEA, but a wide confidence interval, this indicates imprecision and invites to optimize the model. In contrast, a very narrow interval would argue for good precision of the RMSEA in reflecting model fit in the population. AMOS (Arbuckle, 2003) reports a 90% interval around the RMSEA value, supporting the evaluation of model fit (Byrne, 2001).

Results

First, means and standard deviations and zero-order correlations are presented (Table 10 and Table 11).

Table 10

Means and Standard Deviations in Rasch Measures (*n* = 449)

	M	***SD***
SDCinformed	0,35	1,89
SDCconscious	0,69	1,42
SDLC	1,03	1,00
SRLC	1,53	1,86
TSE	0,77	1,08

All scale means are above zero *logits*. As already expected, the highest means are scored in SRLC and SDLC, since the constructs do not yet measure the top end of the continuum. Means in TSE and in both SDC constructs are more modest. The means in SDCinformed are not only lower than in SDCconscious, but also is the standard deviation larger in SDCinformed. The dispersion in SRLC is substantially larger than in SDLC.

Table 11

Zero Order Correlations of the Rasch Measures

	SDCinformed	SDCconscious	SDLC	SRLC	TSE
SDCinformed					
SDCconscious	0,62**				
SDLC	0,09	0,10*			
SRLC	-0,06	-0,04	0,39**		
TSE	0,11*	0,09	0,46**	0,39**	

** Correlation is significant at the .01 level

* Correlation is significant at the .05 level

The most remarkable finding regarding the zero-order correlations of Rasch operationalized interval measures is that there is no significant relationship between SDCinformed and SDLC. In addition, the relationship between SDCconscious and SDLC is weak, though significant. The same goes for the relationships between both SDC constructs and TSE, where only the zero-order correlation with SDCinformed is significant and weak. The zero-order correlation between both SDC constructs is moderately high. Furthermore, SRLC, SDLC and TSE are moderately high correlated.

The Constructs in SEM

Before the constructs were modelled in SEM, we verified whether all possible regressions were linear. Indeed, linear regressions were found to be the best predicting regressions for all hypothesized outcomes. Since there is a lack of research concerning possible causal relations between the five constructs, we had to capitalize on the available literature, such as findings that self-efficacy is a predictor of learning strategies, metacognition, and learning outcomes. However, the question was which metacognition(s) would be predicted by TSE, and whether one metacognition could have a causal path to another. We just had to explore different models, and then select the model with the best fit (Table 12).

Table 12

Modeling the Constructs (n = 449)

	χ^2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null Model	441.00	10	44.10	.00	.73	.59			.31	.29	.34	.00
Model 1	10.36	4	2.59	.04	.99	.98	.99	.97	.06	.02	.11	.30
Model 2	9.80	4	2.45	.04	.99	.97	.99	.97	.06	.01	.10	.34
Model 3	**4.30**	**4**	**1.08**	**.37**	**1.00**	**.99**	**1.00**	**1.00**	**.01**	**.00**	**.07**	**.78**
Model 4	9.64	4	2.41	.05	.99	.97	.99	.97	.06	.01	.10	.35
Model 5	10.29	5	2.06	.07	.99	.97	.99	.98	.05	.00	.09	.46
Model 6	10.29	5	2.06	.07	.99	.97	.99	.98	.05	.00	.09	.46
Model 7	42.24	6	7.04	.00	.97	.92	.92	.86	.12	.09	.15	.00

Model 1: TSE, **SDLC** and SDCconscious as independent variables; SDLC and SDCinformed as dependent variables. A hypothesized regression path from both TSE and SDLC to SRLC, and a hypothesized path from SDCconscious to SDCinformed. A hypothesized regression path from TSE to SDCinformed resulted in a β = .065, p =.12. Therefore that regression path was constrained to zero.

Model 2: TSE, **SRLC**, and SDCconscious as independent variables; SDLC and SDCinformed as dependent variables. A hypothesized regression path from both TSE and SRL to SDLC and a hypothesized path from SDCconscious to SDCinformed. The $\Delta\chi^2$ (0) between model 2 and model 1 was -0.56, ns, so both models were still plausible.

Model 3: TSE, SRLC, and **SDCinformed** as independent variables; SDLC and SDCconscious as dependent variables. A hypothesized regression path from both TSE and SRLC to SDLC and a hypothesized path from SDCinformed to SDCconscious. The $\Delta\chi^2$ (0) as compared with Model 1 is -6.06, and compared with Model 2 $\Delta\chi^2$ (0) = 5.50. Though this difference cannot be calculated since we cannot calculate differences in degrees of freedom, there is a $\Delta\chi^2$ between model 3, versus both model 2 and 1 that can be considered as significant, even thought the models have the same degrees of freedom. Hence, model 3 has the best fit, which is also reflected in the accompanying SEM fit statistics. Also in comparison with the models 4, 5, 6, and 7, model 3 has the best fit.

Model 4: TSE, SDLC, and **SDCinformed** as independent variables; SRLC and SDCconscious as dependent variables. A hypothesized regression path from both TSE and SDLC to SRLC and a hypothesized path from SDCinformed to SDCconscious. A $\Delta\chi^2$ (0) of 5.34 of model 3 as compared with model 4 is a significant difference with similar degrees of freedom, reflecting that the data fit the model better in model 3.

Model 5: A model with TSE and SDCinformed as independent variables, SRLC as a mediating variable, between TSE and SDLC, a hypothesized direct regression path from TSE to SDLC, and a hypothesized path from SDCinformed to SDCconscious. Model 3, as compared with model 5 has $\Delta\chi^2$ (1) = 5.99, p =.01. So, a better fit is achieved in Model 3.

Model 6: A model with TSE and SDCinformed as independent variables, SDLC as a mediating variable, between TSE and SRLC, a hypothesized direct regression path from TSE to SRLC, and a hypothesized path from SDCinformed to SDCconscious. Model 3, as compared with model 6 has $\Delta\chi^2$ (1) = 5.99, p =.01. So, a better fit is achieved in Model 3.

Model 7: A model with TSE and SDCinformed as independent variables. TSE directly predicts both SRL and SDL, and SDCinformed predicts SDCconscious. As compared with model 3 $\Delta\chi^2$ (2) = 37.94, p =.00. hence, model 3 is still the best fitting model.

Based on the above explorations, model 3 was definitively selected as the model with the best fit. Even more important, model 3 was the most plausible model with respect to available theory and research. To verify whether the model would hold in differently composed samples, the total sample (*n* = 449; see Figure 7) was randomly split in a calibration (*n* = 224) and a validation sample (*n* = 225), with the SPSS command 'select cases' and 'random sample of cases'. In the calibration sample the model structure is created, in the validation sample the model structure is controlled for, in other words, validated. After validation the calibration sample and validation sample are combined into a simultaneous sample, which is used for testing the invariance of the regression paths by constraining the relevant parameters to be equal across samples (Byrne, 2001). MSA (multi-sample analysis) is particularly useful to test a pattern of relationships in multiple samples simultaneously. By means of MSA it is possible to investigate to which degree a postulated pattern of relationships is consistent with the observed data in two or more samples (Houkes, 2002). Moreover, MSA provides the possibility to investigate whether a proposed pattern of relationships is invariant (i.e. the structural paths have the same direction and strength) across two or more samples. This procedure provides a powerful validation of the pattern of relationships across samples (Byrne, 2001). Testing for invariance involves specifying a model in which certain parameters are constrained to be equal across the groups under study and then comparing that model with a less restrictive model in which these parameters are free to take on any value (Byrne, 2001). The latter usually is the unrestricted simultaneous model. These nested competing models can be compared by means of the chi-square difference test. A non-significant difference in chi-square indicates invariance (Byrne, 2001).
Table 13, 14, and 15 provide the fit statistics of respectively the calibration sample, the validation sample, the simultaneous sample and the constrained simultaneous sample.

Table 13

Model 3 the Calibration Sample

	χ^2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null Model	223.60	10	22.36	.00	.72	.58			.31	.28	.35	.00
Model cali	1.05	4	0.90	.26	1.00	.99	1.00	1.04	.00	.00	.04	.96

Table 14

Model 3 the Validation Sample

	χ2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*

Null model vali	225.75	10	22.58	.00	.72	.58			.31	.28	.35	.00
Model vali	8.16	4	2.04	.09	.99	.95	.98	.95	.07	.00	.14	.27

Table 15

Model 3 the Simultaneous Calibration and Validation Sample, Unconstrained and Constrained.

	χ2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null model	449.34	20	22.47	.00	.72	.58			.22	.20	.24	.00
Model cali-vali	9.20	8	1.15	.33	.99	.97	1.00	.99	.02	.00	.06	.87
Cali-vali constrained	16.34	13	1.26	.23	.99	.97	.99	.99	.02	.00	.06	.90

The constrained simultaneous model, as compared with the unconstrained model has a $\Delta\chi^2$ (5) = 7.13, p =.21. The non-significance difference in chi square indicates invariance across samples. All parameters could be set equal, except for the correlation between SRLC and SDCinformed that significantly differed in both samples, $\Delta\chi^2$ (1) = 4.65, p =.03. This correlation was the only parameter that was freely estimated in the simultaneous constrained model. In the calibration sample $r_{SRLC\text{-}SDCinformed}$ = -.13, p = .04; in the validation sample $r_{SRLC\text{-}SDCinformed}$ = .02, p = .79.

Figure 7 depicts model 3 in the complete sample (n = 449). The model postulates that both TSE, β = .36, p = .00 and SRLC, β = .25, p = .00 directly predict SDLC, and that SDCinformed directly predicts SDCconscious, β = .62, p = .00. TSE explains 21% in R^2_{SDLC}, and SRLC adds another significant 5% in R^2_{SDLC}. SDCinformed explains 38% in $R^2_{SDCconscious}$.

Model 3 including outcome measures.

For a subsample of respondents in educational sciences and psychology (n = 118) we could obtain educational outcome measures from the administration office, such as grade point average (GPA) in the master trajectory, and achieved ECTS[4] in the master trajectory, providing the means to extend model 3 with educational outcome measures. SEM fit statistics for this extended model are provided in Table 16, and the extended model 3 is depicted in Figure 8.

[4] ECTS = European Credit Transfer System. An agreed upon European standard for study credit points. One ECTS is approximately 28 hours of study load.

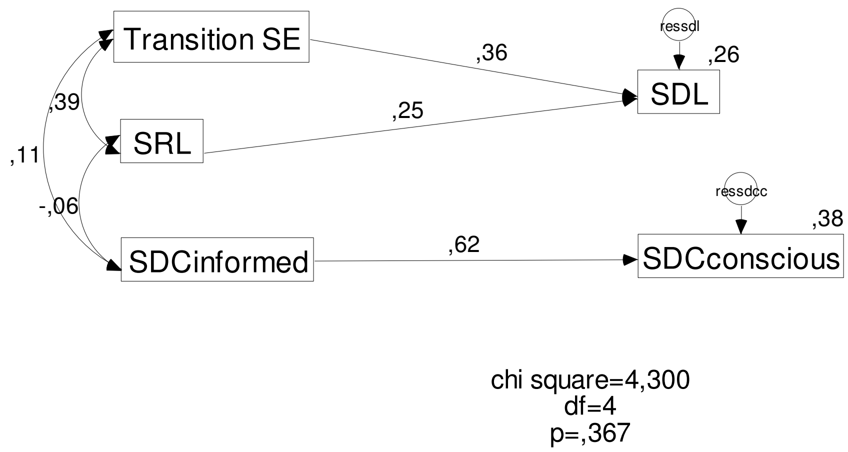

Figure 7. Model 3 (n = 449)

Table 16

Extended Model 3 with Educational Outcome Measures (n = 118)

	χ2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null model 3												
extended	165.36	21	7.87	.00	.72	.63			.24	.21	.28	.00
Model 3												
extended	6.37	13	0.49	.93	.99	.97	1.00	1.07	.00	.00	.03	.98

The extended model 3 postulates that TSE and SRLC indirectly affect GPA via SDLC. TSE, β = .39, p = .00 and SRLC, β = .27 p = .00 simultaneously predict 33% in R^2_{SDLC}. SDLC on its turn predicts 7% of R^2_{GPA}, β = .27 p = .00. Baron and Kenny (1986) have postulated four conditions to specify whether a variable can be considered as a mediator: a) First, the independent and dependent variable should be significantly correlated. In this subsample we found a marginally significant correlation between SRLC and GPA, r =.16, p = .08. We consider an α = .10 as acceptable for a relatively small subsample. In addition, we found a significant correlation between TSE and GPA, r =.23, p = .01. b) There must be a significant relation between the independent variables and the mediator. As demonstrated before, also in this subsample we found a significant correlation between

SRLC and SDLC, r =.46, p = .00, and a significant correlation between TSE and SDLC, r =.52, p = .00. c) the mediator should have significant relationship with the dependent variable when the effects of the independent variable are controlled. When TSE and SRLC were controlled for, the r_{pSDLC_GPA} was .17, p = .07. d) mediation is obtained, when the effects of the independent variables will approximate zero, revealing that all their effects occurred through the mediator. A stepwise regression demonstrated that the regression coefficients of both TSE and SRLC were ns when SDL was included in the second step of the regression.

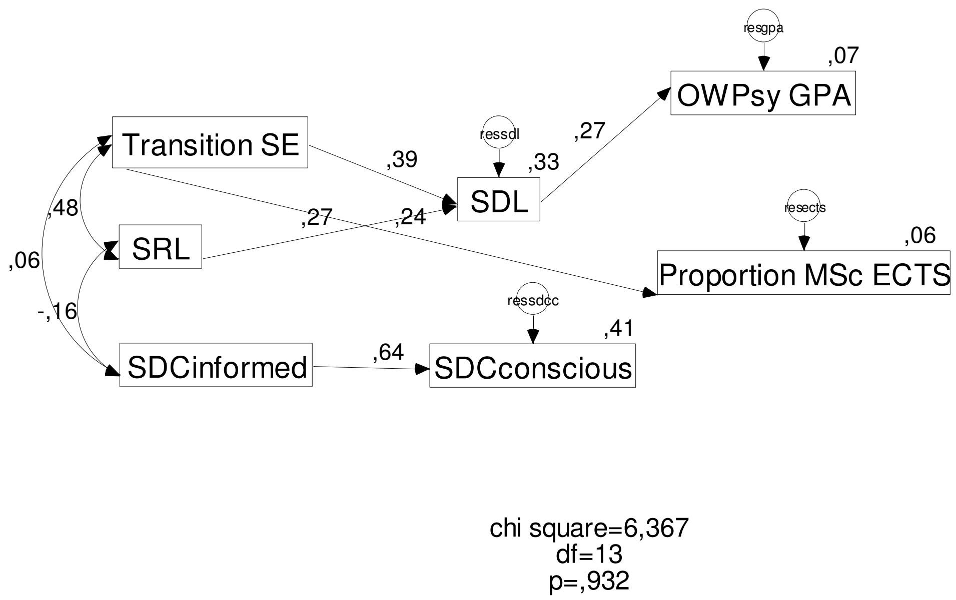

Figure 8. Extended model 3 with educational outcome measures (n = 118)

The model also postulates that TSE directly predicts the proportion of obtained ECTS in the master program, β = .24 p = .01, explaining 6% of $R^2_{proportionMScECTS}$. Thus far, no relations of career constructs with outcomes are identified. Labour market outcome measures, such as mobility and promotion, will be collected in future studies.

Conclusion and Discussion

This methodological study was carried out to answer the central research question: “Can the self-regulating and self-directing learning capabilities (expressed in Rasch interval measures) be combined in a structural model that predicts different learning outcomes?“ This central research question can be confirmed, and the model we developed has demonstrated a good fit.

To answer this central research question we first had to address two preliminary research questions, namely “Can generic self-regulating capabilities and self-directing capabilities be identified as conceptually distinct capabilities?” and “Do the self-regulating and self-directing constructs fit the Rasch rating scale model, to produce Rasch interval measures?”

The first research question could partially be confirmed. The two types of concepts are different in nature, and the distinction between self-efficacy and SRLC is clear. However, arriving at the self-directing concepts diffused the comparisons. Fortunately, the Rasch model, in our second research step, defines generalizable and accurate measures for the theory-based operationalizations of all constructs. After the second step in our research, and assisted by the Rasch model, we could clearly identify five (instead of the proposed four) constructs.

By addressing and answering the third, and central research question, this study has contributed to research and literature, by providing insights in the concepts, and the empirical relationships between self-regulating and self-directing concepts, based on analyses with generalizable and accurate interval measures. Self-regulating learning capabilities (SRLC and TSE) behave as significant predictors of self-directing learning capabilities (SDLC), and the self-directing learning capabilities (SDLC) behave as a mediator between the self-regulating capabilities and educational outcome measures. The relation of SDLC with GPA is an important finding, since meta-analyses have revealed that GPA is a significant predictor of workplace performance (Roth, BeVier, Switzer lll, & Schipmann, 1996). The relations between TSE and SRLC towards SDLC also indicate that SDLC is not merely a maturation like phenomenon, but strongly related to expectancies and metacognitive strategies that are associated with education programs.

Both the factor analyses, CFA analyses, the Rasch model, have clearly identified distinct constructs, thereby finding no evidence for the suggestion of Jossberger at al. (in press) that SRLC might be embedded in SDLC. SRLC also consistently comprised both a task-oriented and a learning trajectory oriented item in the PAF, PCA, CFA, and the Rasch model. Our theory-based and empirically validated conception of SRLC and SDLC shows some resemblance with Pilling-Cormick and Garrison’s (2007) conceptualization, who propose a more inward or outward (meta)cognitive management distinction. However, Pilling-Cormick and Garrison purely focused on a distinction between SRLC and SDLC, while we also considered TSE as a self-regulative construct, just like Pintrich (2004). In addition, we also included self-directed career capabilities, system wise, as self-directing capabilities. While Pilling-Cormick and Garrison (2007) have not carried out empirical studies to study the relations between SRLC and SDLC, we scientifically constructed measures, and empirically verified the relations. Just like Pintrich (2004) we combined all constructs in an interacting, intra-individual system (model). Our conceptualization and empirical findings concerning SRLC are still in line with Pintrich’s (2004) intra-individual system conception, who incorporated motivations, expectancies and cognitions in his intra-individual self-regulating model. A bridge is created between the self-regulating system and a self-directing system. After a reflection on the content validity (Appendix C) the purely metacognitive character of SRLC versus the more strategic and executive metacognitive nature of SDLC the concepts seem to come close to Schön's (1995) concepts of reflection on action and reflection in action.

In all our explorative models self-directing concepts were non significantly related to career concepts, that appear to behave as independent self-directing competences, at least for our groups of high-qualified respondents. This very finding is a contrast with previous research of Raemdonck (2006) who found correlations between SDLC and SDC of $r = .55$ and $r = .71$, amongst both high- and low-qualified employees. Several issues underlie these different research findings. First, we improved the already validated items of Raemdonck (2006) with additional, theory-based items, since we noticed that some self-directing components might have been underrepresented in the instruments as developed by Raemdonck (Appendix 3 reflects on the content validity of the self-regulating and self-directing constructs, and Appendix 4 provides a Rasch analysis of the Raemdonck SDLC scale). The current study shows that predominantly the additional items compose the self-directing constructs, although our literature-based extra items were intended to scaffold the already developed items by Raemdonck (2006), instead of replacing many of them. Using the Rasch model in the current study we were able to identify that specifically the self-directed learning capabilities concept was a problematic concept to be accurately operationalized, as a consequence of the very narrow scale distances and disordered Rasch-Andrich thresholds in the majority of the theory-based items. It is impossible to identify such construct specific measurement problems using CFA procedures that do not provide scale category diagnoses, and that do not separate person abilities and item difficulties on the same scale. Also, CFA does not produce test and sample free estimates that support scale diagnoses. Consequently Raemdonck's (2006) findings were inclined to be strongly affected by item- and group characteristics. However, thanks to Raemdonck's profound conceptual review we could use those fundamentals to improve her previous operationalizations of self-directing learning capabilities. Third, although measurement instruments may be tested for invariance across samples (also our Rasch validated person and item measures are invariant across samples and tests within the intervals of the standard errors), factor correlations can vary across groups (e. g. see our calibration and validation sample, that demonstrated significantly different correlations between SRCL and SDCinformed). The high correlations that Raemdonck (2006) has found in her samples can both be the consequence of the sample and test features, and population characteristics.

The absence of highly significant relations between SDLC and career constructs was not the only remarkable deviation from previous research findings. We also could not identify a significant regression path from TSE to one of the career constructs, as was found by Pinquart et al. (2003). Pinquart et al. did not provide validity data about their self-efficacy construct. They summed up their ten item self-efficacy Likert scale scores, and only communicated a Cronbach alpha of .68. As we have seen in the preliminary analyses, reliability might be a necessary condition for validity, but still is no guarantee for validity. Sometimes scales can appear very 'reliable' without even approximating construct validity. On top of that, structural equation models, like most statistic techniques, assume that the measures that are used are interval measures. When ordinal scales are used in their summed up form, SEM cannot identify the bias that is produced in the analyses as a consequence of fundamentally wrong measurement assumptions (Bond & Fox, 2001, 2007; Byrne, 2001). Hence, our model with Rasch interval measures will more accurately approximate the realistic relationships between TSE and career constructs than the study of Pinquart et al., despite their praiseworthy robust method to identify mediation. Mediation of SDCconscious is still an option, when we can complete our model with labour market outcomes. It might be that SDCconscious then operates as a mediator between SDCinformed and labour market outcomes.

Obviously, our current study has its own shortcomings. In contrast with experimental research, causal models are still no hard evidence of causes and consequences, although structural models are strongly indicative for directions and magnitudes of relationships. Correlational research however can never exclude that possible other important causal factors are overlooked. In addition, and just like most other studies in the field of self-regulating and self-directing constructs, self-report measures are used, which may be affected by social desirable responses or overestimation. Therefore, studies with self-report measures should preferably be triangulated with observational measures, such as content analysis, or measures from different raters, such as supervisors or teachers.

Our explorative structural models have suggested that SRLC was a determinant of SDLC in revealing better fits of models that expressed such a relationship. However, alternative models do not exclude that this relationship is the other way around, or reciprocal. Conceptually, it would be logical when SRLC precedes SDLC. Yet, using the same logic we can also argue that SDLC is a predictor of SRLC. We believe that in practice both concepts will continuously interact, and that the role of SRLC is to deeply reflect *on learning actions* that have already occurred, and to improve personal learning capabilities. In contrast, SDLC is more a reflection *in action* construct, combining all components during task execution (See Appendix C). As compared with SRLC, SDLC makes the impression of a more productivity-oriented construct, while SRLC purely represents metacognitive, deeper reflection that precedes, adjusts, and innovates new actions. The nature of SDLC makes the impression of being more instantaneous, and when we combine this with our empirical measurement findings of collapsing categories and disordered Rasch-Andrich thresholds, we now suppose that SDLC's empirical measurement structure might be partly dichotomous, partly rating scale, instead of a 100% rating scale construction (cf. Linacre, 2010; Stone, 1998). SDLC might combine features of being apparent versus absent, and features that express a certain intensity. Follow-up studies should clarify this proposition.

Finally, we still have not been able to develop a SDLC or a SRLC construct that measures the higher end of the continuum of the underlying variable. Further scale development of both constructs seems to be an important subject for future research.

References

Arbuckle, J. L. (2003). *Update to the AMOS User's Guide*. Chicago: SmallWaters Corporation.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Beltyukova, S. A., Stone, G. E., & Fox, C. M. (2004). Equating student satisfaction measures. *Journal of Applied Measurement, 5*(1), 62-69.

Bijker, M. M., Wynants, G., & Van Buuren, J. A. (2006). *A comparative study of the effects of motivational and affective factors on studying statistics.* Paper presented at the ICOTS7, Salvador, Brazil. http://www.stat.auckland.ac.nz/~iase/publications/17/5H2_BIJK.pdf

Blom, S., Severiens, S., Broekkamp, H., & Hoek, D. (2005). *Zelfstandig leren van allochtone en autochtone kinderen in het Studiehuis*. [Independent learning among immigrant and native children in the "Study house"]. Amsterdam: Instituut voor de Lerarenopleiding.

Bode, R. K. (2001). Partial credit model and pivot anchoring. *Journal of Applied Measurement, 2*(1), 78-95.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. New Jersey: Lawrence Erlbaum Associates.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model. Fundamental measurement in the human sciences* (Second ed.). Mahwah, New Jersey: Lawrence Erlbaum.

Brockett, R.G., & Hiemstra, R. (1991). *Self-direction in adult learning: Perspectives on theory, research, and practise.* London: Routledge.

Byrne, B. M. (2001). *Structural Equation Modeling with Amos. Basic concepts, applications and programming*. New Jersey: Lawrence Erlbaum Associates.

Chen, C. S. (2002). Self-regulated learning strategies and achievement in an introduction to information systems course. *Information Technology. Learning, and Performance Journal, 20.*(1), 11-25.

Claes, R., & Ruiz-Quintanilla, S. A. (1998). Influences of early career experiences, occupational group, and national culture on proactive career behavior. *Journal of Vocational Behavior, 52*, 357-378.

Curda, L. K. (1997). *Validation of a measure of teacher's self-efficacy and outcome expectations in the content domains of reading and mathematics.* Doctoral thesis, University of Oklahoma, Norman.

Eccles, J. S., Wigfield, A., Flanagan, C. A., Miller, C., Reuman, D., & Yee, D. (1989). Self-concepts, domain values, and self-esteem: relations and changes at early adolescence. *Journal of Personality, 57*(2).

Fouarge, D., De Grip, A., & Nelen, A. (2009). *Leren en werken* (No. ROA-R-2009/3). Maastricht, the Netherlands: Research Centre for Education and the Labour Market.

Forsyth, R., Saisangjan, U., & Gilmer, J. (1981). Some empirical results related to the robustness of the Rasch model. *Applied Psychological Measurement, 5*, 175-186.

Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology, 45*(1), 30-45.

French, J.R.P. Jr, Rogers, W., & Cobb, S. (1981). A model of person-environment fit. In: L. Levi (Ed.), *Society, Stress and Disease* (Volume 4, pp 39-44). New York: Oxford University Press.

Guglielmino, L. M. (1977). *Development of the self-directed learning readiness scale.* Unpublished doctoral thesis, University of Georgia, Georgia.

Guglielmino, L. M., Long, H. B., & Hiemstra, R. (2004). Self-direction in learning in the United States. *International Journal of Self-Directed Learning, 1*, 1-17.

Hiemstra, R. (2000). Self-directed learning: the personal responsibility model. In G. Straka (Ed.), *Conceptions of self-directed learning* (pp. 59-74). Münster: Waxmann.

Houkes, I. (2002). *Work and individual determinants of intrinsic work motivation, emotional exhaustion and turnover intention.* Doctoral thesis, Maastricht University, the Netherlands.

Jackson, J. W. (2002). Enhancing self-efficacy and learning performance. *The Journal of Experimental . Education, 70*(3), 243-254.

Jossberger, H., Brand-Gruwel, S., Boshuizen, H. P. A., & Van de Wiel, M. W. J. (in press). The challenge of self-directed and self-regulated learning in vocational education: A theoretical analysis and synthesis of requirements. *Journal of Vocational Education and Training*.

King, Z. (2004). Career self-management: Its nature, causes and consequences. *Journal of Vocational Behavior, 65*, 112-133.

Knowles, M (1970). The modern practice of adult education. New York: *Association Press*

Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers*. Chicago: Follett.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Linacre, J. M. (2009a). Winsteps® (Version 3.69.0). [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved December 29, 2009. Available from http://www.winsteps.com/.

Linacre, J. M. (2009b). *A User's Guide to Winsteps Ministeps Rasch Model Computer Programs*. Retrieved December 28, 2009, from www.winsteps.com/winpass.htm

Linacre, J. M. (2010). *Transitional categories and usefully disordered thresholds*. Retrieved from http://www.oerj.org/View?action=viewPDF&paper=2

Linnenbrink, E. A., & Pintrich, P. R. (2002). Achievement goal theory and affect: an asymmetrical bidirectional model. *Educational Psychologist, 37*(2), 69-78.

Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly*(19), 119-137.

Marsh, W., Roche, L. A., Pajares, F., & Miller, D. (1997). Item-specific efficacy judgments in mathematical problem solving: the downside of standing too close to trees in a forest.

Contemporary Educational Psychology(22), 363-377.

Mellenbergh, G. J. (1980). Theorie op verschillende niveaus. [Theories at the different levels]. *Nederlands tijdschrift voor de psychologie*(35), 275 - 288.

Mellenbergh, G. J. (1996). Measurement precision in test Score and item response models. *Psychological Methods, 1*(3), 293-299.

Messick, S. (1995). Validity of psychological assessment. Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Mirvis, P.H., & Hall, D.T. (1994). Psychological Success and the Boundaryless Career. *Journal of Organizational Behavior, 15*, 365-380.

Näring, G. W. B., Hoogduin, K. A. L., & Keijser, C. M. P. (2004). A Rasch analysis of the Stanford hypnotic susceptibility scale, form c. *The International Journal of Clinical and Experimental Hypnosis, 52*(3), 250-259.

Nieuwenhuis, L., Gielen, P., & Nijman, D. J. (2008). *Leven lang leren voor vitaliteit*. [Lifelong learning for vitality]. Retrieved April 8, 2010, from http://cinop.brengtlerentotleven.nl/downloads/publicaties/A00467.pdf

Owen, T. R. (2002). *Self-directed learning in adulthood: A literature review* (ERIC document No. ED461050).

Parker, P., Hall, D. T., & Kram, K. E. (2008). Peer coaching: A relational process for accelerating career learning. *Academy of Management Learning & Education, 7*(4), 487-503.

Pilling-Cormick, J., & Garrison, D. R. (2007). Self-directed and self-regulated learning: Conceptual Links. *Canadian Journal of University Continuing Education, 33*(2), 13-33.

Pinquart, M., Juang, L. P., & Silbereisen, R. K. (2003). Self-efficacy and successful school-to-work transition: A longitudinal study. *Journal of Vocational Behavior, 63*, 329-346.

Pintrich, P.R. (2000). Educational Psychology at the millennium: a look back and a look forward. *Educational Psychologist, 34*(4), 221 - 226.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-407.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40.

Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in Education* (2 ed.). New Jersey: Merrill Prentice Hall.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ).*Unpublished manuscript, Ann Arbor – Michigan: University of Michigan.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and*

Psychological Measurement, 53, 801-813.

Raemdonck, I. (2006). *Self-directedness in learning and career processes. A study in lower-qualified employees in Flanders.* Doctoral dissertation, University of Gent, Belgium.

Roth, P. L., BeVier, C. A., Switzer lll, F. S., & Schipmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology, 81*(5), 548-556.

Schön, D. A. (1995). The new scholarship requires a new epistemology. *Change, 27*(6), 6-35.

Semeijn, J. H., Van der Velden, R. K. W., Heijke, H., Van der Vleuten, C., & Boshuizen, H. P. A. (2005). *Competence indicators in academic education and early labour market success of graduates in health sciences* (No. ROA-RM-2005/2E). Maastricht, the Netherlands: Research Centre for Education and the Labour market.

Smith, E. V., & Dupeyrat, C. (2001). Toward establishing a unified metric for performance and learning goal orientations. *Journal of Applied Measurement, 2*(4), 312-336.

Stone, M. H. (1998). Rating scale categories: Dichotomy, double dichotomy, and the number two. *Popular Measurement, Spring 1998*, 61-6

Taherbhai, H. M., & Young, M. J. (2004). Pre-equating: a simulation study based on a large scale assessment model. *Journal of Applied Measurement, 5*(3), 301-318.

Trochim, W. M. K. (2002). *Measurement Validity Types*. Retrieved from http://www.socialresearchmethods.net/kb/measval.htm

Van Buuren, J. A. (2008). *Van vakgericht naar competentiegericht statistiekonderwijs.* [From conventional to competency-based statistics education].Doctoral thesis. Open University of the Netherlands, Heerlen.

Van der Heijden, B. (2002). Individual career initiatives and their influence upon professional expertise development throughout the career. *International Journal of Training and Development, 6*(2), 54-79.

Van der Klink, M., & Boon, J. (2002). The investigation of competencies within professional domains. *Human Resource Development International, 5*(4), 411-424.

Wild, K.-P., & Schiefele, U. (1993). Induktiv versus deduktiv entwickelte Fragebogenverfahren zur Erfassung von Merkmalen des Lernverhaltens. [Inductively versus deductively developed scales to measure learning strategies]. *Unterrichtswissenschaft* (21), 312-326.

Wolfe, E. W. (2000). Equating and item banking with the Rasch model. *Journal of Applied Measurement, 1*(4), 409-434.

Wolters, C. A., Pintrich, P. R., & Karabenick, S. A. (2003). *Assessing academic self-regulated learning.* Paper presented at the Conference on Indicators of Positive Development: Definitions, Measures, and Prospective Validity.

Wright, B. D., & Mok, M. M. C. (2004). Introduction to Rasch Measurement. In B. D. Wright & M. H. Stone (Eds.), *Making Measures*. Chicago: Phaneron Press.

Wright, B. D., & Stone, M. H. (1999). *Measurement essentials* (2nd ed.). Wilmington, Delaware:

Wide Range.

Yu, C. H., & Osborn Popp, S. E. (2005). Test equating by common items and common subjects: concepts and applications. *Practical Assessment, Research & Evaluation, 10*(4), 1-19.

Appendix A

Table A1

Means and Standard Deviations of Items in the Validated Scales (Sources: Pintrich et al., 1991,1993; Raemdonck, 2006, and Bijker et al., 2010).

Itemnr	Scales and items	*M*	*SD*	Pi	Ra	Bij
	SDCinformed					
43	I keep myself well informed about opportunities to develop my career.	3,24	0,97		1	
45	I know whom to approach to find a proper job.	3,04	0,94			1
53	When I am searching for a job, I know which companies offer the best career opportunities in my field of study.	3,09	0,96			1
54	I know exactly what I want to achieve in my career during the following years.	3,23	0,97		1	
58	I always keep myself posted about what is important in the eyes of employers.	3,21	0,90			1
60	For me it is clear what I want to achieve in my career.	3,47	0,91		1	
	SDCconscious					
46	I think it is important to have ambitious career goals.	3,50	0,95			1
47	I regularly express my career interests to people who can be of importance for my career.	3,35	1,00		1	
52	I think it is relevant to get in touch with people who can be of importance to my career as much as possible.	3,73	0,87		1	
55	I always think hard about which steps I have to take to achieve my career objectives.	3,50	0,82		1	
56	I keep my network informed about what I want to achieve in my career.	2,82	0,96			1
59	I use my network to increase the chances to get my desired job.	3,32	0,99			1
61	I consult others to get a realistic picture about opportunities to increase my personal effectiveness.	3,53	0,87			1
63	My career trajectory will be a central issue in the negotiations with my future employer.	3,04	0,98			1
66	I check whether my career goals are realistic.	3,75	0,66			1
	SDLC					

9	I constantly gather information about courses to stretch my competencies.	3,35	0,99			1
12	I have clear ideas about what and how I want to learn.	3,85	0,79			1
16	If I notice that a certain learning strategy does not work I change my approach.	3,87	0,65			1
18	I try to get involved in projects that offer me a lot of learning opportunities.	3,82	0,76		1	
23	I aim to achieve the top level in my profession.	3,35	1,02			1
25	I am committed to continuously increase my competencies.	4,19	0,59			1
27	I am well informed about excellent providers of education programs in my field.	3,17	0,94			1
32	I want to perform well, even if I do not like the learning task or theory.	3,82	0,74	1		
	SRLC					
15r	I often find learning so boring that I quit before I finish what I planned to do.	4,08	0,85	1		
22r	Frequently I do not identify learning moments, because I am thinking about other things.	3,48	0,87	1		
26	When I participate in an education program I make sure that I complete that program.	4,11	0,81	1		
28	Even when a learning task is dull, I keep working until I have finished the task.	3,93	0,73	1		
38	I never give up when I am learning something difficult.	3,75	0,77		1	
	TSE					
69	I know for sure that I will be able to flexibly apply theory in business practice.	3,81	0,70			1
72	Compared with other students in this master program, I expect to do well.	3,24	0,82	1		
74r	My performance in this master program is modest in comparison with others.	3,02	0,86	1		
76r	I am worried about my study results in this master program.	2,36	0,99	1		
79r	When I supervise others I am troubled about how poorly I am doing.	2,26	0,82			1
80	I am sure that I have the potential to become a real business leader.	3,69	0,78			1
82	I am certain I can understand even the most difficult tasks and topics.	3,50	0,89	1		
84	Compared with other students in this master program I think that I have a clear view of business.	3,43	0,75			1

Appendix B

CFA Fit Statistics (after Rasch modelling the scales)

Table B1

CFA Stastistics SDCinformed (n = 449)

	χ^2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null Model SDCinformed	1035.56	15	69.04	.00	.47	.26			.39	.37	.41	.00
Model SDCinformed	11.73	8	1.47	.16	.99	.98	1.00	.99	.03	.00	.07	.75
Model Raemdonck	*207.1*	*77*	*2.69*	*.00*	*.93*	*.90*	*.93*		*.07*	*.06*	*.08*	

Table B2

CFA Stastistics SDCconscious (n = 449)

	χ^2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null Model SDCconscious	1178.97	36	32.75	.00	.48	.37			.27	.25	.28	.00
Model SDC conscious	42.70	23	1.86	.01	.98	.96	.98	.97	.04	.02	.06	.76
Model Raemdonck2	*207.1*	*77*	*2.69*	*.00*	*.93*		*.90*	*.93*	*.07*	*.06*	*.08*	

Table B3

CFA Stastistics SDC one dimension (n = 449)

	χ^2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null Model SDC 1 dimension	2579.07	105	24.56	.00	.36	.27			.23	.22	.24	.00
Model SDC 1 dimension	258.16	81	3.19	.00	.93	.89	.93	.91	.07	.06	.08	.00
Model Raemdonck2	*207.1*	*77*	*2.69*	*.00*	*.93*		*.90*	*.93*	*.07*	*.06*	*.08*	

Table B4

CFA Stastistics SDLC (n = 449)

	χ^2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null model												
SDLC	355.69	28	12.70	.00	.77	.71			.16	.15	.18	.00
Model												
SDLC	30.44	19	1.60	.05	.98	.97	.97	.95	.04	.01	.06	.81
Model												
Raemdonck1	212.8	104	2.05	.00	.93	.90	.90		.05	.04	.06	
Model												
Raemdonck2	125.8	77	1.63	.00	.95		.95		.04	.03	.05	

Table B5

CFA Stastistics SRLC (n = 449)

	χ^2	*df*	*Cmin/df*	*P*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null												
Model												
SRLC	488.58	10	48.86	.00	.63	.45			.33	.30	.35	.00
Model												
SRLC	4.07	4	1.02	.40	1.00	.99	1.00	1.00	.01	.00	.07	.80

Table B6

CFA Stastistics TSE (n = 449)

	χ^2	*df*	*Cmin/df*	*P*	*RMR*	*GFI*	*AGFI*	*CFI*	*TLI*	*RMSEA*	*LO*	*HI*	*PCLOSE*
Null													
model													
TSE	723.03	28	25.82	.00	.17	.65	.55			.24	.22	.25	.00
Model													
LESE	38.14	13	2.93	.00	.03	.98	.94	.96	.92	.07	.04	.09	.13

Appendix C

Content validy: A reflection on the conceptual framework

This section is a reflection on the extent to which hypothesised characteristics are represented in the constructs.

Table C1

SRLC (based on the Motivated Strategies for Learning framework; Pintrich et al., 1991, 1993).

Itemnr	**Loadings mbk et al.**	**Loadings Raemdonck**	**Categorization**	**Item**	**Rasch scale position**
<u>*Goal setting*</u>	Is conceptualized by Pintrich et al. in motivation and expectation constructs.				
<u>*Choosing strategy*</u>	Is conceptualized by Pintrich et al. in learning strategies constructs.				
<u>*Execute strategy*</u>	Is conceptualized by Pintrich et al. in learning strategies constructs.				
<u>*Monitor and evaluate*</u>					
28	.75		Adjust goals, plan, implementation	Even when a learning task is dull, I keep working until I have finished the task.	Moderately easy item to endorse, within – 1 SD from average person measure
26	.63		Reflect on the self as a learner	When I participate in an education program I make sure that I complete that program.	Easiest item to endorse, positioned at – 1 SD from average person measure
15r	.52		Overcome complexity and	I often find learning so	Moderately difficult,

			negative emotions	boring that I quit before I finish what I planned to do. (reversed)	within -1 SD from average person mean
38	.67	.48/.54 (in SDL)	Overcome complexity and negative emotions	I never give up when I am learning something difficult.	Positioned at the average person mean level.
22r	.39		Evaluate impact of strategy (result)	Frequently I do not identify learning moments, because I am thinking about other things. (reversed)	Most difficult item to endorse. Within +1 SD range of average person measure

"Evaluate impact of strategy" and "overcome complexity and negative emotions" are both hypothesized SDLC monitoring and evaluation facets that did not fit the SDL model, but are a substantial part of the SRLC-model
Not represented[5] (neither in SRL nor SDL): "Prioritize learning" and "Register progress"

The SRLC construct is exclusively composed of monitoring en evaluating items. Pintrich et al. consider the items above as regulating, monitoring, and evaluating, expressed in effort, persistence, perseverance, and focus. Focus is operationalized in negatively formulated items that have shown to be very robust, both in the CFA, and in the Rasch rating scale analyses.

For a fourth construct in a PAF or PCA it was remarkable that the item loadings were that high (except for 22r). With some fantasy, the 'near goal characteristic', as described by Jossberger et al. (submitted) might be identified in item 28. In contrast, item 26 represents a distal goal, implicating that both near and distal goals are part of SRLC. Regulation, persistence and focus in directing the learning strategies are the core characteristics of the items that fit both the factor analyses, and the Rasch rating scale model. In contrast with the theory, overcoming complexity or negative affect, and evaluating the impact of a strategy are exclusive for SRLC, and consistently did not fit the SDLC model, neither in factor analyses, CFA, nor the Rasch model.

The purely regulating, monitoring and reflection characteristic can underlie the fact that the SRLC construct is a separate phenomenon, not being encompassed in SDLC. Category distances in the SRLC

[5] Because the items in these subcategories did not load in factor analyses and did not fit the Rasch model.

construct are much wider than the very small categories that are characteristic for the SDLC construct. SRLC has the content characteristics of internalized critical thinking (about the self, and individual learning behaviour).

Table C2

SDLC, Based on the Theoretical Framework of Raemdonck (2006)

Itemnr	**FL_mbk**	**FL_ir**	**Face valid subcat**	**Item**	**Comments**
Goal setting					
23 Moderately difficult. Within range - 1SD of average person measure	.57		Anticipate future learning needs.	I aim to achieve the top level in my profession.	Mbk et al.
25 Easily endorsed item; 1SD < 2SD from average person measure	.57		Formulate learning goals	I am committed to continuously increase my competencies.	Mbk_item
Not represented: Detect knowledge skills or gaps* Diagnose personal learning needs* ** = not fitting the Rasch model					
Choosing strategy					
18 Easy item, > 1SD < 2SD from average person measure	.45	.49/.46	Select appropriate strategy	I try to get involved in projects that offer me a lot of learning opportunities.	IR_item **un**modified (item25 in SDL)
12 Easy item, > 1SD < 2SD from average person measure	.40		Develop a learning plan	I have clear ideas about what and how I want to learn.	Mbk et al.

person measure				

Not represented:
Collect information about learning opportunities*
Identify human and material resources for learning*
* = not fitting the Rasch model

Execute strategy

27 Moderately difficult item in the Rasch model (below average person measure). Within range -1SD of average person measure	.45	Explore learning market and work environment.	I am well informed about excellent providers of education programs in my field.	Mbk et al.
9 Moderately difficult item. Within range -1SD of average person measure	.42	Explore learning market and work environment.	I regularly look for information about courses that can support me in further developing my competencies.	Mbk et al.

Not represented:
Express learning interests*
Networking to create learning opportunities*
Ask advice to realize learning plan*
* = not fitting the Rasch model

Monitor and evaluate

32 Easy item, >1SD < 2SD from average person measure	.27	Reflect on the self as a learner	I want to perform well, even if I do not like the learning task or theory.	Pintrich et al.
16 Easy item, > 1SD < 2SD from average	.35	Adjust goals, plan, implementation.	If I notice that a certain learning strategy does not work I	Mbk et al.

person measure	change my approach.
Not represented:	
Prioritize learning*	
Register progress*	
Overcome complexity and negative emotions**	
Evaluate impact of strategy**	
* = not fitting the Rasch model	
**Represented in SRL	

The two higher loading items are in the category goal setting. Yet, the two most difficult items, according to the Rasch rating scale model, are about being well informed about the learning market (item 27, an executive item), and item 9, another executive item, namely on collecting information about courses. The construct is well balanced, regarding the four expected components.
Goal setting: 2 items (25%)
Choose strategy: 2 item (25%)
Execute strategy: 2 items (25%); including most difficult items
Monitor and evaluate: 2 items. (25%)
However, that feature can also be interpreted as a compromise.

In Raemdonck's (2006) study the highest loading items were: "I keep my manager informed on what I want to reach in my career", which was not adopted in our education context (either school to work, or work to school). The other higher loading item in her study was "I never give up when I am learning something difficult", which has been incorporated in SRLC (see above). Instead, an original Pintrich et al. item has been incorporated in SDLC (item 32) in the evaluation section.

Table C3

SDCinformed Based on the Theoretical Framework of Raemdonck (2006)

Itemnr	**FL_mbk**	**FL_ir**	**Face valid subcat**	**Item**	**Comments**
			SDC informed		
Goal setting					
53 Moderately difficult item, within -1 Sd of average	.77		Detect career opportunities	When I start searching for a job, I know which companies offer the best career opportunities in my field of study.	Mbk et al.

person measure.					
54 Moderately difficult item, within -1 Sd of average person measure	.60	.61	Formulate learning goals	I know exactly what I want to achieve in my career during the following years.	IR_item modified (or nr2 in SDC)
60 The easiest item to endorse. Within -1 Sd of average person measure.	.53	.54	Diagnose personal career expectations	For me it is clear what I want to achieve in my career.	IR_item modified (or nr10 in SDC)

In principle all subcategories are represented. However, “anticipate future labour market developments” has been replaced by “anticipate shared values” (such as achievement orientation) in SDC_conscious. That subcategory has been introduced based on the person-environment-fit model or Michigan model (French, Rogers, & Cobb, 1981).

Choosing strategy

45 Most difficult item, within +1 SD of average person mean.	.74		Select appropriate strategy	I know whom to approach to find a proper job.	Mbk et al.
43 Moderately difficult	.70	.64	Collect information about job opportunities	I keep myself well informed about opportunities to	IR_item modified (nr12 in SDC)

item, within -1 Sd of average person measure			develop my career.	

Not represented:
Develop a career plan*
Identify key persons is represented in SDC_conscious.
* = not fitting the Rasch rating scale.

Execute strategy

58 Moderately difficult item, within -1 Sd of average person measure	.62	Explore learning market and work environment	I always keep myself posted about what is important in the eyes of employers.	Mbk et al.

All subcategories are represented (Express career interests, networking and self presentation, and asking advise to realize the career plan are represented in SDC_conscious).

Monitor and evaluate

Adjust goals, plan, implementation is represented in SDC_conscious

Not represented:
Reflect on strengths, preferences, capacities*
Prioritise career* (also loading insufficiently in Raemdonck's CFA analysis)
Overcome complexity and negative emotions* (also insufficient loading in Raemdonck's CFA analysis)
Evaluate the impact of the strategy * (also insufficient loadings in Raemdonck's CFA analysis)
Evaluate course of career * (*did* load sufficiently in Raemdonck's CFA analysis)
*= Not fitting the Rasch rating scale or partial credit model (disordered average measures or step calibrations)

Table C4
SDCconscious
Goal setting

46 Moderately difficult item, within – 1 Sd from average person measure	.52		Anticipate shared values, such as achievement motivation.	I think it is important to have ambitious career goals.	Mbk et al.
55 Relatively easy item to endorse. At -1 SD from average person measure	.47	.67	Formulate career goals	I always think hard about what I have to undertake to achieve my career objectives.	IR (mod) 24E
<u>*Choosing strategy*</u>					
52 Easy to endorse item, > 1SD < 2sd from average person measure	.61	.56	Identify key persons	I think it is relevant to get in touch with people who can be of importance to my career as much as possible.	IR_item modified (nr 27in SDC)
<u>*Execute strategy*</u>					
56 Most difficult item.	.61		Express career interests	I continuously keep my network informed about what I want to achieve in my career.	Mbk et al.

Within + 1Sd of average person measure.					
63 Second most difficult item, within + 1 SD from average person measure	.55	(.47/.57)	Explore learning market and work environment	My career trajectory will be a central issue in the negotiations with my employer.	Mbk et al. (drastic modification of item 29 IR "I find it important to negotiate with my manager about new steps in my career")
59 Moderately difficult item, within – 1SD from the person mean	.78		Ask advice to realize career plan	I use my network to increase the chances to get my desired job.	Mbk et al.
47 Moderately difficult item, within – 1SD from the person mean	.74	.72	Networking and self presentation	I regularly express my career interests to people who can be of importance for my career.	IR (mod) 12E
61 Moderately difficult item,	.53		Ask advice to realize career plan	I consult others to get a realistic picture about opportunities to increase my personal	Mbk et al.

within – 1SD from the person mean			effectiveness.	
Monitor and evaluate				
66 Most easy to endorse. > 1SD < 2SD from average person measure	.46	Adjust goals, plan, implementation	I check whether my career goals are realistic.	Mbk et al.

SDCinformed.

Goal setting: 3 items (50%)

Chose strategy: 2 items (33%); includes most difficult item

Execute strategy: 1 items (17%)

Monitor and evaluate: 0 items. (0%)

SDCconscious.

Goal setting: 2 items (22%);

Chose strategy: 1 item (11%)

Execute strategy: 5 items (56%); includes the two most difficult items.

Monitor and evaluate: 1 item. (11%)

In SDCinformed the emphasis is on goal setting. The construct does not include any monitoring or evaluation item. SDCconscious is dominated by execution items that also include the most difficult items. Both constructs do not or hardly consist of monitoring and evaluation items. All theoretically hypothesized SDC monitoring and evaluation subcategories have shown to be absent in the construct(s), both in the analyses of Raemdonck, and in the current study, except for 'adjusting goals, plan, or implementation'. The career processes constructs are the least characterized by reflection. The emphasis in career processes is on a) execution and b) goal setting.

Conclusion

Moving from SRLC downwards to the SDC career constructs there seem to be different levels of reflection represented in the constructs, with SRLC at the top, followed on quite some distance by SDLC, and only very modestly represented in one SDC construct. Specifically in comparison with SRLC, both SDLC and the career constructs are characterized by a strong emphasis on execution, and goal setting. The hypothesized components are all present in all constructs. Subcategories, that are not present in the constructs, have not survived the factor analyses, CFA, and/or the Rasch model.

Appendix D

Raemdonck's (2006) operationalization of SDLC[6]

Table D1

The Raemdonck SDLC Construct in the Rasch model (in the educational science and psychology sample[7]; n = 371)

Item	Infit	Outfit	Measure	Error	PTMEA	Miscellaneous	
8	1.09	1.11	1.60	.06	.46	Item 8 is measuring outside the range of fitting items (see person and item maps). obscuring that the construct is measuring below the average person measure.	
19	1.00	1.03	.72	.06	.46		
11	.98	1.05	.43	.06	.49		
5	1.07	1.09	.26	.06	.40		
37	.94	.89	-.05	.09	.45		
7	1.03	1.06	-.14	.07	.40		
18	.91	.90	-.22	.08	.46		
2	1.02	1.01	-.29	.08	.40	Misfits in item 2 and item 34	
10	.94	.97	-.34	.07	.43		
21	.79	.78	-.34	.08	.57	Disordered Rasch-Andrich thresholds in seven items, being item 2, item 10, item 11, item 21, item 34, item 36, and item 37.	
36	1.07	1.09	-.35	.07	.37		
39	1.19	1.22	-.42	.07	.30		
34	.99	.92	-.85	.09	.37		
All items							
Mean	1.00	1.01	.00	.07		Person Reliability	.58
SD.	.10	.11	.60	.01		Person Separation	1.17
All persons						Item Reliability	.98
Mean	1.02	1.01	.84	.42		Item Separation	8.08
SD	.62	.62	.67	.10		Cronbach alpha	.67

The average item infit and outfit mean square seem to reasonably fit the model, however the person infit and outfit mean squares show some underfit. The average item measures seem to be in ascending order from -0.85 logit in item 34 to 1.60 logit in item 8. However, item 8 is measuring outside the aimed for item continuum in items (> + 2SD; see the person and item map). Just like the whole pool of items that was designed for this construct, more than 50% suffered from disordered Rasch-Andrich

[6] Raemdonck's original item 17 was omitted, since the experts did not consider the item as suitable for the current sample. Skipping an item has no consequences for the position of other items or persons on the scale.
[7] In the business administration sample the majority of these items was already omitted.

thresholds, indicating that at least one of the item's scale categories was not at one moment in time the most probable category that the person scores identified.

```
OBSERVED AVERAGE MEASURES FOR PERSON (scored) (ILLUSTRATED BY AN OBSERVED
CATEGORY)
-2        -1         0         1         2         3         4
|---------+---------+---------+---------+---------+---------|  NUM   ITEM
|                        12    3    4           5           |    4  Q8
|                                                           |
|                                                           |
|                    1  2  3  4            5                |    8  Q19
|                                                           |
|                   1    2  3 4       5                     |    6  Q11
|                   1  2    34      5                       |    2  Q5
|                                                           |
|            1      2  3   4       5                        |   12  Q37
|              1      2   3 4     5                         |    3  Q7
|         1        2     3  4     5                         |    7  Q18
|                     2  13  4     5                         |    1  Q2
|           1         2   3 4    5                          |    5  Q10
|             1  2     3   4       5                        |    9  Q21
|                    1 2 3 4     5                          |   11  Q36
|                 1      23 4    5                          |   13  Q39
|                                                           |
|                                                           |
|            1         3 24    5                            |   10  Q34
|---------+---------+---------+---------+---------+---------|  NUM   ITEM
-2        -1         0         1         2         3         4

                    1122 234 23 2 1 2 1
       1    23 4154665645 771 75 1 8 2 1 7 5  4 1      1    2  PERSON
            T       S       M      S       T
       0           10 30 40 60 80  90         99               PERCENTILE
```

Figure D1. The small categories in SDL items.

Figure D1 demonstrates how narrow the categories of the SDLC items are. Instead of a common category range between -4.0 to +4.0 logits, the categories in this 'scale' do not measure below -1.0 and beyond +2 *logits* (when we skip item 8, which does not match the content continuum of the underlying variable). A same, but far more extended version of such a scale (all with fitting in- and outfits) could be produced with the complete pool of empowering SDL items, however with a similar percentage of misfits and threshold problems. When item categories are that small, thresholds get distorted, biasing the measurement. In Figure D1, only misfits can be identified, but not threshold problems.

Table D2 and Figure D2 demonstrate an example of a threshold problem, such as in item 21, "When I want to learn something that can be useful for my job I take the initiative." It demonstrates that category 4 dominates and oppresses the categories 2 and 3.

Table D2

Disordered Rasch-Andrich Thresholds in Item 21

CATEGORY LABEL	STRUCTURE MEASURE	S.E.	SCORE-TO-MEASURE AT CAT.	----ZONE----		50% CUM. PROBABLTY	COHERENCE M->C	C->M	RMSR	ESTIM DISCR	
1	NONE		(-2.68)	-INF	-2.14		0%	0%	2.1731		1
2	-1.03	.60	-1.58	-2.14	-1.19	-1.72	0%	0%	1.5365	1.38	2
3	-1.21	.32	-.81	-1.19	-.32	-1.19	50%	14%	.8898	1.21	3
4	-1.28	.16	.61	-.32	2.23	-.66	69%	98%	.2017	1.07	4
5	2.14	.13	(3.27)	2.23	+INF	2.17	82%	12%	.7422	1.26	5

M->C = Does Measure imply Category?
C->M = Does Category imply Measure?

In the structure measure column it is obvious that the threshold calibration measures for both category 3 and category 4 are descending instead of ascending and that only the threshold measure of category 5 is higher than the threshold measure of category 2.

```
        CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
R  1.0 +                                                       +
O      |11111                                                  |
B      |     111                                           5555|
A      |        11                                       55    |
B   .8 +          1                                    55      +
I      |           1                                 55        |
L      |            1              4444444          5          |
I      |             1           44       44       5           |
T   .6 +              1         4           4    55            +
Y      |               1       4             44 5              |
    .5 +                1     4                *               +
O      |                1     4               5 4              |
F   .4 +                 1   4               5   4             +
       |                  1 4              55      44          |
R      |                  14              5          4         |
E      |              22222*33333        5            44       |
S   .2 +            22   3* *2   33    55              44      +
P      |         222   334   12    3355                  44    |
O      |     2222    33 4     122 555333                   444 |
N      |22222    3333444      5***222   333333                4|
S   .0 +*************555555555    111**************************+
E      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
       -4    -3    -2    -1     0     1     2     3     4     5
        PERSON [MINUS] ITEM MEASURE
```

Figure D2. Disordered Rasch-Andrich thresholds in item 21.

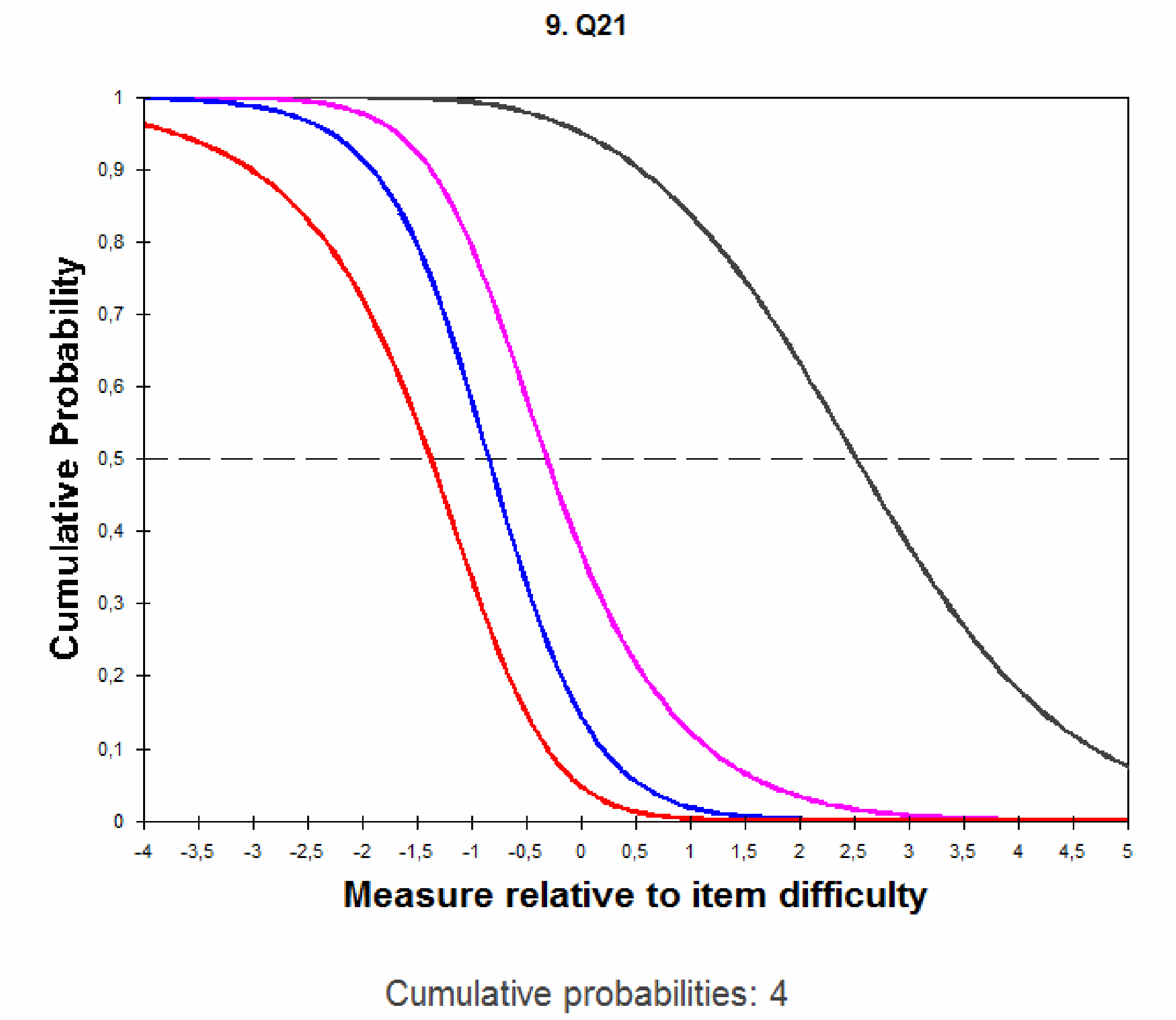

Figure D3. The narrowness of the Scale Categories 2 and 3, Relative to Category 4 in item 21, expressed in Rasch-Thurstone[8] Thresholds (Linacre, 2010).

The scale has an item separation of 8.08, and an item reliability .98. The person separation of 1.17 is insufficient, just like the person reliability of .58. It implicates that the scale seems to identify 11.11 distinct item difficulty strata that the participants distinguish and 1.89 distinct SDL-capability strata, distinguished by the items. The item difficulty strata are confounded. They are overestimated as a consequence of the most difficult item 8, that is measuring outside the continuum of items, which indicates that item 8 is inclined to measure a different dimension. The person capability strata are insufficient to be useful in distinguishing individuals with high and low SDL-abilities. The range of item calibrations is from -0.85 (item 34) l*ogit* (item 8) to 1.60 logit. The ordering of these items seems

[8] Each Rasch-Thurstone threshold is the point on the latent variable where the cumulative category probabilities for categories below the threshold are 0.5.

to provide some evidence of the construct validity, given that the items seem to be logically ordered from least to most difficult. However the most difficult item is measuring a different dimension, and the easiest item has distorted Rasch-Andrich thresholds. The person and items map of Raemdonck's SDLC is depicted in Figure 4. The variance that is explained by the measures is 39.6%; 13.1% is explained by the persons, and 26.4% by the items.

The average item measure is .00 l*ogit* (*SD* .60) and the average person measure is 0.84 *logit* (*SD* 0.67). The most difficult item measure (item 8, 1.60 l*ogit*) is measuring far beyond (+0.76 *logit*) the average person measure. In fact, the item (a negotiation item) is beyond the acceptable item difficulty range of + 2SD (see Figure 4). The most difficult item is 0.88 *logits* removed from the second most difficult item, and comprises 36% of the whole range in measures.. The SDL scale is confounded by measurement problems, and not suited for group comparisons, because measures with threshold problems and misfits are unstable, and bias statistical analyses, particularly in situations in which the qualitatively ordered Likert edition of the scale is summed up. In addition, the scale is not sufficiently discriminative to identify persons with low and high SDL-capabilities. The argument that the above scale is validated in a sample of high-qualified respondents and therefore will not hold in a sample of low-qualified employees is not plausible, since Rasch item calibrations are invariant within standard error estimates (Fox & Jones, 1998). This can also be noticed in Figure D4 that demonstrates that Raemdonck's SDLC-scale is far too easy to endorse for our population, and that all items (except item 8) measure below the average person measure.

A secondary analysis of Raemdonck's data in the Rasch model is an option, since one item (item 18) in the scale above overlaps the Rasch validated SDLC scale in our study. Hence, Raemdonck's SDLC data could be anchored on the item measure of item 18 of our SDLC-scale.

Our previous analyses of the SDLC scale with the whole range of 35 items already identified that it is likely that at least some theory-based SDL items will have the characteristics of a yes/no scale. Dichotomous scales need a relatively large range of items achieve discriminative power. In addition, dichotomous items are not suited for CFA analyses, since CFA measurement models need scales that have at least three categories (Byrne, 2001).

```
--------------------------------------------------------------------------------

           PERSON - MAP - ITEM
               <more>|<rare>
    4             .  +
                     |
                     |
                     |
                  .  |
                     |
                     |
                     |
    3                +
                  .  |
                     |
                     |
                  #  |
                 .#  |
                     |
                 .# T|
    2                +
                .##  |
                     |
             .#####  |  Q8
              .#### S|
                     |
             .#####  |T
          .########  |
    1        .######  +
        .########## M|
         .#########  |  Q19
                     |S
            .######  |
            .######  |  Q11
         ##########  |  Q5
               .### S|
    0            .#  +M Q37
                 .#  |  Q7
                  #  |  Q18    Q2
                 .#  |  Q10    Q21    Q36    Q39
                  # T|
                  .  |S
                     |
                  .  |  Q34
   -1                +
                     |
                  .  |T
                     |
                     |
                     |
                     |
                     |
   -2                +
               <less>|<frequ>
 EACH "#" IS 4. EACH "." IS 1 TO 3
```

Figure D4. Person and item map of the SDLC-scale of Raemdonck (n = 371; graduates in educational science and psychology)