Carlo F. Dondena Centre for Research on Social Dynamics

DONDENA WORKING PAPERS

A scale for the measurement of self-efficacy for learning (SEL) at university

Jane E. Klobas, Stefano Renzi, Maria Luisa Nigrelli

Working Paper No. 2
URL: www.dondena.unibocconi.it/wp2

September 2007

Carlo F. Dondena Centre for Research on Social Dynamics
Università Bocconi, via Guglielmo Röntgen 1, 20136 Milan, Italy
http://www.dondena.unibocconi.it

ISSN: 2035-2034

A scale for the measurement of self-efficacy for learning (SEL) at university

Jane E. Klobas
Carlo F. Dondena Centre for Research on Social Dynamics
Università Bocconi
via Guglielmo Röntgen 1
20136 Milan
Italy
and
UWA Business School
University of Western Australia
PO Box 1164
Nedlands 6909
Western Australia
jane.klobas@unibocconi.it, jane.klobas@uwa.edu.au

Stefano Renzi
Carlo F. Dondena Centre for Research on Social Dynamics
Università Bocconi
via Guglielmo Röntgen 1
20136 Milan
Italy
stefano.renzi@unibocconi.it

Maria Luisa Nigrelli
TILS Spa
Viale dell'Oceano Pacifico, 11
00144 Roma
Italy
marialuisa.nigrelli@tils.com

Abstract

Students' experiences at university prepare them for a future in which they are expected to engage in life-long learning. Self-efficacy theory suggests that a person's beliefs in their capacity to learn will influence their participation in learning. This paper describes development of a new scale to measure self-efficacy for learning (SEL) among university students, designed to be appropriate for both campus-based and online learning, and for administration in a battery of tests on student development. Undergraduate students (*n* = 265) in a business school in Milan and a department of psychology in Rome participated in the final study. Beginning with a random sample of 200 participants, item response theory and exploratory factor analysis with LISREL were used to identify a 10 item scale to measure SEL. The scale's properties were confirmed in a second random sample of 200 participants, using confirmatory factor analysis in LISREL. Correlation with expected grades was, consistent with earlier studies, moderately small (.22), but statistically significant.

Keywords

academic self-efficacy, self-efficacy for learning, SEL, learning, university, measurement scale

Introduction

There is increasing interest in the metacognitions that arise from participation in education. Such metacognitions include understanding and control of processes that contribute to learning – such as planning, monitoring and evaluating – as well as the development of attitudes and beliefs, such as self-efficacy (SE) beliefs, that contribute to self-regulation and control of learning processes (Sternberg, 1998; Zimmerman, 1995). Among older learners, including university students, these meta-responses to learning are believed to contribute, not just to learning performance, but also to their continued participation in learning, as life-long learners (Klobas, Renzi, Francescato, & Renzi, 2002). In this paper, we concentrate on SE, in the context of study of meta-response to participation in learning at university. Our goal is to develop a scale to measure SE for learning (SEL) that can be used in situations where SE forms one of several meta-responses of interest. Thus, our goal is to develop a parsimonious scale capable of discriminating among different levels of SEL.

Self-efficacy refers to "the judgments of one's capability to organize and execute the courses of action required to produce given attainments" (Bandura, 1997, p. 3). It varies along three dimensions: level, strength, and generality (Bandura, 1997; Holladay & Quinones, 2003). People vary in the *level* of difficulty of the tasks that they believe they can perform, and in the *strength* of their beliefs about their ability to attain a given level of difficulty. Thus, we can ask students how confident they are (strength of SE) that they can understand either half or all (two different levels) of the concepts described in a lecture. *Generality* refers to the notion that "efficacy beliefs associated with one activity can be generalized to similar ones within the same activity domain or across a range of activities" (Holladay & Quinones, 2003, p. 1094). Thus, we can argue that there exists a general form of SEL that applies to a range of activities in the domain of learning at university. It is this general form of SEL that we are interested in measuring, rather than SE for performance in specific subjects, because our interest is in a general form of SE.

What is the mechanism by which SEL attained at university might contribute to participation in life-long learning? Self-efficacy has been found to contribute to persistence at learning tasks, and to expected and actual performance of learning tasks (Zimmerman, 1995). The measured effects of generalized academic SE on expected and actual performance are small, particularly when student ability and other influences on performance are taken into account (Valentine, DuBois and Cooper, 2004, in a meta-analysis, estimated a mean effect on actual performance of .08 although correlations of up to .36 are reported in the literature), but the effects are nonetheless statistically significant and observed consistently across different samples (Valentine, DuBois, & Cooper, 2004; Wood & Locke, 1987; Zimmerman, 1995; Zimmerman, Bandura, & Martinez-Pons, 1992). Bandura (1997) proposes that SE also affects choice behaviours. This is particularly important if we want to encourage life-long learning. People are more likely to perform tasks for which they have high SE and avoid those for which their SE is low. We propose that SEL developed at university contributes to SEL for participation in life-long learning.

Measuring SE in academic settings

Most research on SE in academic settings has concentrated on the SE of school students, and most of this research uses the academic SE scale developed by Bandura (Bandura, 2001; Pintrich & De Groot, 1990; Valentine et al., 2004; Zimmerman, 1995). Bandura's academic SE scale is quite a lengthy instrument, measuring SE for performance of activities associated with various academic subjects (e.g., arithmetic) as well as activities associated with general performance at school (e.g. organizing, making friends). The instrument is not readily adapted to study SE outside the school domain.

Several academic SE scales have been developed for use with post-secondary students (Chemers, Garcia, & Hu, 2001; Elias & Loomis, 2000; Owen & Froman, 1988; Wood & Locke, 1987). There is no dominant scale as there is for study of SE among school students, and the scales that have been developed vary in the extent to which they actually address SE or related concepts such as outcome expectancy.

The scale that comes closest to meeting the objectives of our current study is the academic SE scale developed by Wood and Locke (1987). This scale was also used by Maurer and Pierce (1998). The scale considers six activities that are performed by university students in their studies in general (class concentration, memorization, understanding, explaining concepts, discriminating concepts, and note.taking), rather than focusing on specific subjects. In their development of this scale, Wood and Locke paid careful attention to measurement of both level and strength of efficacy beliefs. Level was taken into account by presenting students with gradations of challenge in percentage terms, e.g. "Memorize 60% [70%, 80%, 90%, 100%] of facts & concepts". Thus, the Wood and Locke scale addresses level, strength and generality in ways that make it an appropriate base for our study. On the other hand, the scale was developed in the mid-1980's and includes activities, such as class concentration, which are not necessarily key activities in courses that incorporate online learning. Our current study therefore sought to develop a new scale, based on the Wood and Locke scale, that would be appropriate for use with students who study online as well as in the classroom.

Study I

Materials and Methods

Measurement

The scale used in study 1 was adapted from the scale used by Wood and Locke (1987) to measure university students' academic SE. For study 1, we removed two activities from the scale: note-taking and class concentration, activities whose role is ambiguous in computer-supported learning environments. We added three activities: ability to organize work to meet course deadlines, connecting ideas and updating knowledge. The activity descriptions are given, in English, in appendix 1.

Although the Wood and Locke (1987) scale presented students with gradations of challenge in percentage terms, only those levels of performance that had sufficient variation to discriminate between respondents were used to measure SE. The lowest, and sometimes the highest, levels of performance were removed for measurement. In developing the current scale, we took into account recent research on SE scales and Bandura's guidelines for scale construction (Bandura, 2001; Maurer & Pierce, 1998; Pajares, Hartley, & Valiante, 2001; Smith, Wakely, De Kruif, & Swartz, 2003). We presented gradations of challenge in the form *most* and *all* or *usually* and *always* rather than as percentages (e.g. "Soon after a lesson is over, I am able to remember *most* of the concepts"; "Soon after a lesson is over, I am able to remember *all* of the concepts"). The lower gradation represented a *relative* level of performance and the higher an *absolute* level of performance.

All items were measured on an 11 point scale ranging from 0 (*I am definitely not able to do this*) to 10 (*I definitely can do this*). Maurer and Pierce (1998) describe a study by Mudgett and Quinones in which such a scale was more effective than a Likert-type scale when gradations of challenge were presented in more generic terms than percentages. The use of the 11 point scale is also supported by Pajares, Hartley and Valiante's (2001) findings that greater scale variation provides more satisfactory results when measuring SE. An extract from the scale, translated into English, appears in Figure 1.

Think about your current activities as a student. Read each of the following statements **carefully**, then circle the number that best represents your response, where:
0 indicates: *I am* ***definitely not able*** *to do this*
10 indicates: *I* ***can definitely*** *do this*

Circle the appropriate number:

1	I am able to organize my activities so that I can meet *most* course deadlines	0--1--2--3--4--5--6--7--8--9--10
2	I am able to organize my activities so that I can meet *all* course deadlines.	0--1--2--3--4--5--6--7--8--9--10
3	Soon after the end of a lesson, I am able to remember *most* of the key concepts.	0--1--2--3--4--5--6--7--8--9--10
4	Soon after the end of a lesson, I am able to remember *all* of the key concepts.	0--1--2--3--4--5--6--7--8--9--10
5	I can understand *most* of the key concepts covered in my course.	0--1--2--3--4--5--6--7--8--9--10
6	I can understand *all* of the key concepts covered in my course.	0--1--2--3--4--5--6--7--8--9--10

Figure 1. Extract from study 1 SE for learning scale showing measurement scale

An item is a useful indicator of SE only if it can discriminate among different levels of confidence in performing the behaviour of interest (Bandura, 1977). The scale contained several items which were not expected to discriminate, but which acted as anchors for other items. For example, participants' confidence that they could recall *all* of the key concepts soon after the end of a lesson (item 4 in Figure 1) was expected to be quite varied. The anchor for this item asked participants to report confidence in their ability to recall *most* key concepts. Relatively little variation was expected on this item, and we did not intend to use it in measuring SE. This is consistent with Wood and Locke's (Wood & Locke, 1987) elimination of non-discriminating levels of challenge from their measurement scale.

Participants

The participants were 1737 Italian university students, 394 studying business in Milan and 1343 studying psychology in Rome. Of these students, 282 were in their final year of study (112 in Milan and 170 in Rome) and the rest were in their first year of study (282 in Milan and 1173 in Rome).

Procedure

The scale was administered both at the beginning of semester and following examination, in two successive years. In the first year, 564 students participated in the study, while in the second year, there were 1173 participants. Descriptive statistics were used to identify non-discriminating items, which were recognized by high mean scores and high skew. Cronbach α was estimated for the remaining items. The validity of the scale was examined in tests of changes in SE between the two administrations using ANOVA. The data gathered in the first year of the study tested the hypothesis that SEL increased among final year students as they gained mastery of the subject they were studying but decreased among first year students who are likely to underestimate the difficulty of study at university. In the second year of the study, we hypothesized that SEL would increase for final year students participating in a course that involved online collaborative learning, but be static for final year students participating in a course with no collaborative learning component.

Results

Ten (10) discriminating items were identified. The discriminating items formed a scale with high internal reliability in all administrations. Cronbach α ranged from .89 to .91.

The scale was able to discriminate between changes in SEL among students in different years of their course. In the first year of the study, SE increased among final year students (from $M = 6.55$, $SD = 1.35$ to $M = 6.96$, $SD = 1.33$), and decreased among first year students (pre-course $M = 7.03$, $SD = 1.38$, post-course $M = 6.56$, $SD = 1.26$, interaction $F = 33.03$, $DF = 1, 192$, $p < .001$). In the second year, SEL again decreased among first year students (pre-course $M = 6.73$, $SD = 1.26$, post-course $M = 6.45$, $SD = 1.36$, $F = 21.78$, $DF = 1,374$, $p < .001$). It increased significantly for final year students in the course with an online collaborative learning component (pre-course $M = 6.74$, $SD = 1.52$, post-course $M = 7.11$, $SD = 1.29$) but not for students taking a wholly classroom-based course (pre-course $M = 6.81$, $SD = 1.5$, post-course $M = 6.7$, $SD = 1.35$, interaction $F = 7.15$, $DF = 1, 154$, $p < .01$).

Discussion

Despite the success of the scale in identifying changes in SEL in different learning situations, we believed some improvements could be made. Scores on the scale were relatively high and changes in SE relatively small, even if statistically significant. Was it possible to develop a scale that had a greater degree of discrimination by adding items of greater difficulty?

Item response theory (IRT) was used to study if the scale could be improved by the addition of some more difficult items. Because all response categories (from 0 to 10) were present for all but three of the items, and these items had no response for just one category, we used a polychotomous rating scale Rasch model (Bond & Fox, 2001). All pre-course scores gathered during study 1 were combined for this analysis. The file contained 1330 complete cases. (Cases with missing values and multivariate outliers were removed.) The analysis was conducted using RUMM (www.rummlab.com.au).

The scale fit the Rasch model well. Reliability, as measured by the RUMM separation index, was .89. Total item-trait interaction χ^2 was 158.52, $DF = 90$, $p = .000012$. Because χ^2 is sensitive to large sample sizes, RUMM allows χ^2 to be adjusted by nominating a reduced sample size for the goodness of fit test. With an adjusted sample size of 500, all items demonstrated good fit with $p > .05$.

Despite the satisfactory fit, the results of the Rasch analysis demonstrated that the distribution of item difficulties could be improved. Figure 2 shows the distribution of items by difficulty, in the lower part of the chart. The upper part of the chart shows the distribution of respondent scores. Together, these distributions indicate that there is little variation in the difficulty of the items (all clustered around the standardized mean difficulty of 0.0) and that the majority of respondents scored above the standardized midpoint of the scale.

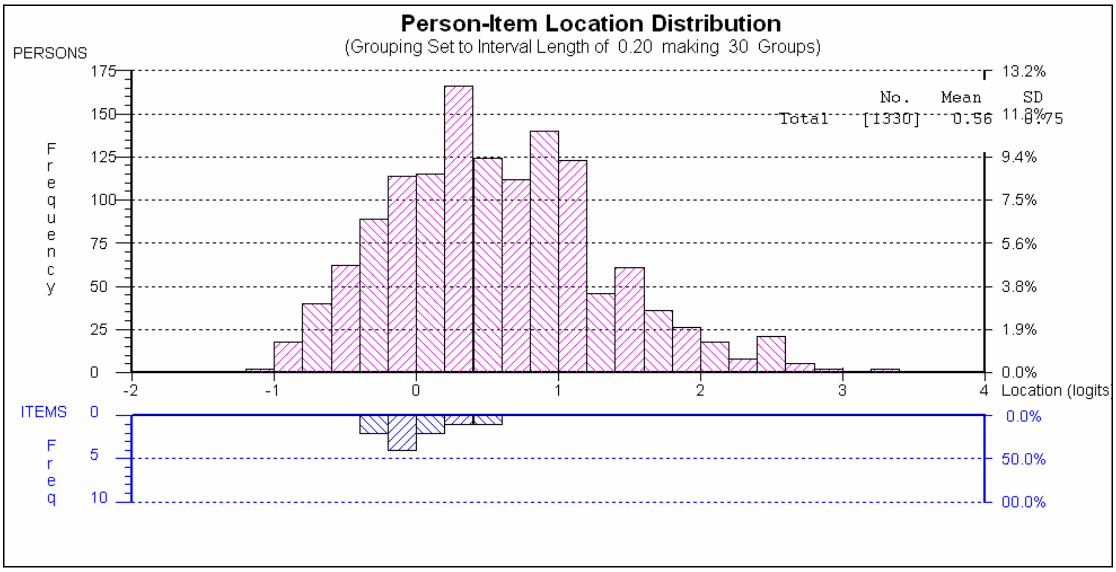

Figure 2. Item-person locations for initial 10 item scale

Study II

Materials and Methods

Measurement

In order to provide a more sound theoretical base to the scale, and to potentially identify some more difficult items, we prepared an extended set of items based on a combination of Bloom's (1969) *Taxonomy* of educational objectives and the activities that students perform as they identify information from sources outside the classroom. Bloom distinguished between six levels of educational achievement: knowledge (observation and recall of information; ability to list, define, describe, quote, name people and facts), comprehension (understanding), application, analysis (seeing patterns; identifying and organizing elements of knowledge), synthesis (generalizing; relating knowledge from several areas; using old ideas to create new ones; predicting; drawing conclusions), and evaluation (comparison and discrimination; assessing theories and ideas; making reasoned choices) (University of Victoria Counselling Services Learning Skills Program, 2003). We were able to match two of the study 1 item sets to the lowest level of the taxonomy, and four to the next four levels, but there were no items to measure SE for activities directed toward learning at the highest level (evaluation). We developed three new items as potential items to measure evaluation in a more difficult scale. In addition, the existing items focused on activities usually associated with classroom-based learning, so we added some candidate items at the lower levels to address self-study of learning materials and learning from other students.

The items in the study 1 scale were primarily concerned with the cognitions or information processing activities associated with learning. Only one item addressed the self-directed gathering of new information from sources such as the library and the World Wide Web. A further set of candidate items was developed for activities associated with gathering information.

After pilot testing, 16 candidate items were added to the scale: three to measure evaluation, six to measure self-study and learning from other students at lower levels of Bloom's taxonomy, and seven to measure activities associated with information gathering. The survey form that contained the full set of items from which the new scale would be developed contained 46 items, 22 of them absolute and 24 relative. The items appear, in English translation, in appendix 2. They are available in Italian from the authors.

Participants

The participants were 265 Italian university students in various years of study, 105 business students in Milan and 160 psychology students in Rome. There were 107 males and 152 females (6 did not provide information on gender).

Procedure

The students completed the survey in the middle of a semester, after they had received some feedback on their academic performance. The relative items were treated as response anchors, and only the absolute items were included in analysis.

Missing value analysis identified one item with more than 2% missing values. There was also a high proportion of extreme scores on this item. The item referred to the hypothetical situation of participation in an online discussion forum. Because this is not a common learning situation, the item was removed from further analysis.

Two samples of 200 were drawn at random from the file of 265 cases. Two exploratory analyses were conducted on the first sample: a Rasch analysis in RUMM, and an exploratory factor analysis using LISREL. The item response analysis identified items that fit the Rasch model based on both student SEL score and item difficulty. The LISREL analysis enabled identification of a parsimonious set of items that reflected the theory on which the scale was based. Confirmatory factor analysis on the model that resulted from the exploratory factor analysis was conducted (again using LISREL) with the second sample. Finally, the range of difficulties of items in the confirmed model was examined using Rasch analysis with the full sample of 265 cases. (Missing values were replaced with means for the LISREL analyses.)

As a test of concurrent validity, scores on the new scale were correlated with students' expected future performance. Given the results described in the introduction to this paper, a moderately small but statistically significant correlation was expected. For this test, students were asked to report the score that they expected in their next exam and the total marks out of which the exam would be scored. We transformed these data into a percentage expected score.

Results

Initial Rasch analysis

The Rasch model was applied to the 21 potential items using RUMM. Four items were removed because of poor fit. The final set of 17 items exhibited good fit to the Rasch model (item-trait interaction $\chi^2 = 44.85$, $DF = 34$, $p = .1$). Reliability was high as measured by the RUMM separation index (.91) and Cronbach alpha (.91). Item difficulty was evenly distributed around the mid-point of the scale (Figure 3). Mean scores on this scale were lower (.32 compared with .56) with a higher standard error (.04 compared with .02) than on the study 1 scale, indicating that a scale based on these 17 items should have better discriminatory ability than the study 1 scale.

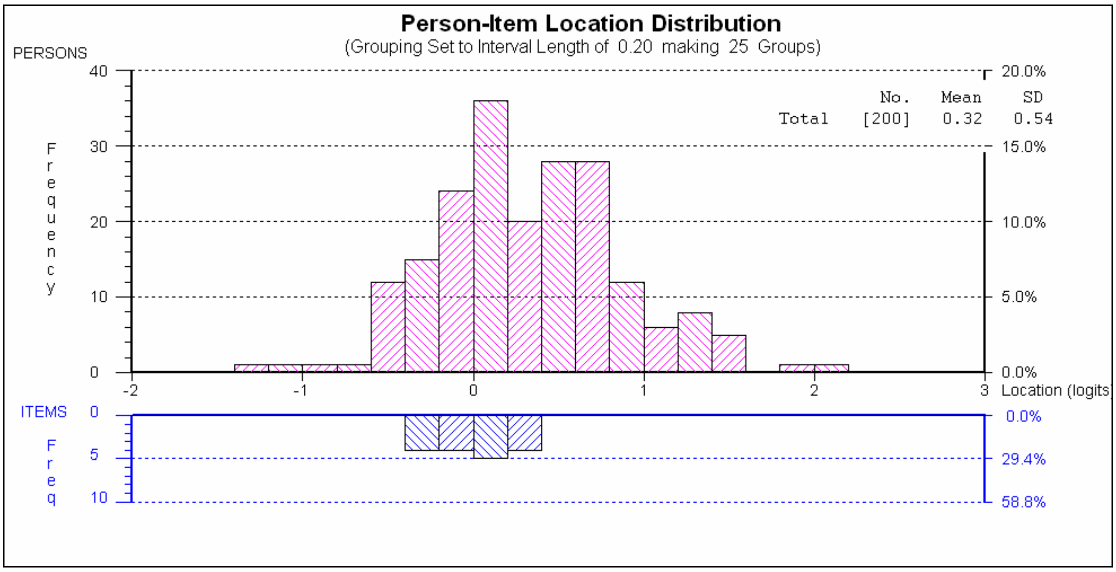

Figure 3. Person-item location plot for new 18 item scale

Exploratory factor analysis with LISREL

An initial factor model was built in LISREL for the Rasch model solution. Because the items were derived from two theoretical bases, Bloom's *Taxonomy* and student activities in finding resources for their study, the model contained two factors which we labelled Info Processing and Finding, respectively. The factors were expected to be distinguishable, but highly correlated, given that the full set of items met the Rasch model criteria for a uni-dimensional scale. The initial model included all 18 item that emerged from the Rasch analysis. The expected loading of the items on each factor is given in appendix 2. Although all paths in this model were significant, the fit was poor ($\chi^2 = 323.37$, $DF = 118$, $p < .001$, GFI = .84, AGFI = .79, RMSEA = .09, p(RMSEA < .05) < .01).

Modification indexes indicated that the model could be improved by the addition of correlations between variables. Most of the pairs of correlated variables contained variables that measured learning at the same level of Bloom's taxonomy. Rather than adding paths for these inter-item correlations, we removed items with the goal of obtaining a parsimonious scale that included only one item for each level of the taxonomy. Items were progressively removed from the model, based on their content. In general, the items that were removed had more complex wording or referred to more complex situations than the items that were retained. Six items were removed on this basis. Although this process resulted in a model with good fit, there remained one pair of items to measure one level of Bloom's taxonomy. We removed one of the items.

The resulting 10 item two factor model (Figure 4) had good fit (χ^2 = 44.99, *DF* = 34, *p* = .1, GFI = .96, AGFI = .93, RMSEA = .04, *p*(RMSEA < .05) = .68). All paths were statistically significant. Individual item R^2 ranged from .33 to .65. The factors (sub-scales) were correlated .68. Cronbach α for the full scale was .86. For the Info Processing subscale, it was .83 and for the Finding subscale .81.

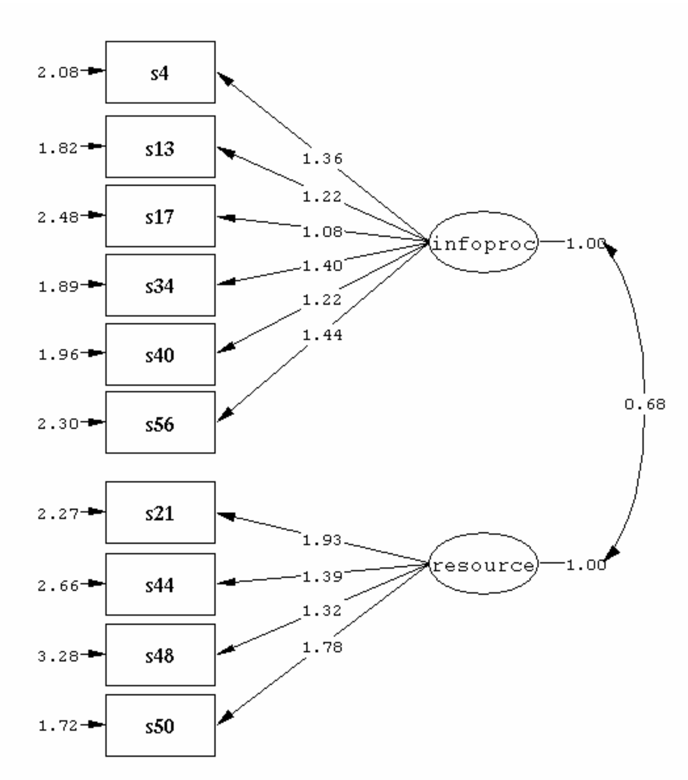

Figure 4. Model resulting from exploratory factor analysis with LISREL

Confirmatory factor analysis with validation sample

The model that resulted from the exploratory factor analysis was tested with the second sample of 200 cases taken at random from the original file of 265. The 10 item two factor structure was confirmed. Fit was good (χ^2 = 38, *DF* = 34, *p* = .29, GFI = .96, AGFI = .94, RMSEA = .02, *p*(RMSEA < .05) = .87). All paths were statistically significant. Individual item R-squared ranged from .32 to .67. The factors (sub-scales) were correlated .68. Cronbach alpha for the full scale was 84. Alpha for both the Info Processing and Finding subscales was .8.

Rasch analysis of item difficulty

RUMM was used to check the distribution of item difficulties in the new 10 item scale. The full 265 cases were used for this analysis. The items were equally distributed either side of the mid-point of the item difficulty scale. Individual scores had a lower mean (.26) and a wider distribution (*SD* = .52, *SE* = .03) than in study 1 where the original 10 item scale was used. The item-person location plot for the new 10 item scale appears in Figure 5.

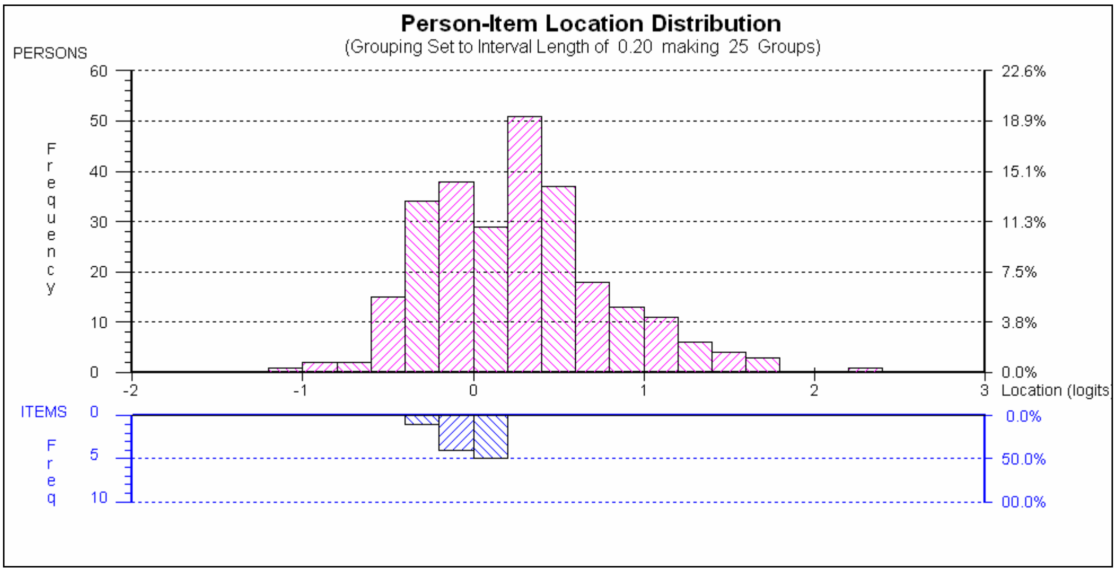

Figure 5. Item-person locations for new 10 item scale

Test of concurrent validity

The correlation between SEL as measured on the new scale and expected score on next assessment was .22 (n = 164, $p < .01$), a small but significant correlation. The magnitude of this correlation is similar to the mean correlation of .27 found by Wood and Locke (1987), and higher than average loading of .08 estimated by Valentine et al. (2004).

Discussion

This research succeeded in its goal to develop a theory-based SE for learning (SEL) at university scale that incorporated items across a range of difficulties. The scale contains items to measure SE for cognitive tasks at each of the levels in Bloom's (1969) taxonomy of educational objectives. It also contains items to measure students' SE for finding information from the primary physical and electronic resources available to them, the library and the world wide web. The scale is capable of discriminating among university students. Furthermore, the new 10-item scale is short enough to be used in batteries of tests, such as tests of students' meta-responses to their learning at university.

The new scale exhibited similar concurrent validity to the Wood and Locke (1987) scale on which it was based, i.e. a moderately small but statistically significant correlation with students' expectations of future exam performance. Wood and Locke provided three explanations for the modest correlation of SE with performance and expected performance at university. Firstly, studies that obtain higher correlations between SE and performance are usually laboratory studies in which a specific task is performed within minutes of completing the SE assessment whereas the time between SE assessment and completion of an assessment at university is a matter of months. Secondly, the range of tasks required for success at university is complex while feedback is given only for overall performance, thus students have less information on which to base their assessment of SE. Thirdly, they note that students already at university are already pre-selected for their ability to perform at university, reducing the variation in SE for learning at university available in a sample of university students. These explanations also apply to the results of the current study. Wood and Locke conclude that, given this situation, the significant correlation of SE with expected performance is “impressive”. While we might not go so far, we believe that identification of a mechanism by which students’ performance at university can be improved, regardless of ability, is important.

The methodology used in this study merits some comment. We have mixed methods (some would say, philosophies) of scale construction to arrive at a scale that satisfies both classical and item response theory. IRT enabled us not only to examine the difficulty of the items on the scale, it provided evidence of the ability of the scale to discriminate among students. It also enabled a preliminary cull of items to those that met the criteria of the Rasch model. It would have been difficult, however, to obtain the final 10-item scale with IRT alone. The Rasch model assumes a uni-dimensional scale and does not aid in identification of potential sub-scales. If we had stopped after the initial Rasch analysis in Study 2, we would have identified a uni-dimensional scale that exhibited satisfactory qualities from the point of view of IRT and the traditional Cronbach α measure of reliability. That would have been satisfactory, but by proceeding with further analysis we were able to identify a parsimonious set of items that still meets our goals in terms of item difficulty and theoretical robustness. The end result is a parsimonious scale that contains items that meet the demands of both classical and item response theory. We therefore expect it to be robust in application with new samples of university students.

The small sample size in study 2 does, however, suggest caution in using this scale without further validation. Because there were only 265 participants, the samples used in the exploratory and confirmatory stages of study 2 contained quite a lot of overlap. Future research would usefully validate both the full 10-item scale and the existence of the two sub-scales.

In our introduction, we discussed the potential contribution of SEL to participation in life-long learning. But, we have not tested this relationship. Future research would fruitfully study students as they move from university into the workplace to determine if high SEL (appropriately measured for the workplace) is indeed associated with making choices to participate in further learning opportunities, persistence during participation in these learning opportunities, and performance as a result of participation.

Acknowledgements

The studies in which the data for this paper were gathered were supported by the Italian Ministry for Instruction, University and Research and Bocconi University. We thank the teachers of the courses in which the surveys were administered.

References

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Reviews, 84*(2), 191-215.

Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. New York: W H Freeman.

Bandura, A. (2001). *Guide for Constructing Self-Efficacy Scales (revised)*, Retrieved 9 May 2005, from http://www.emory.edu/EDUCATION/mfp/Bandura/

Bloom, B. S. (1969). *Taxonomy of Educational Objectives: The Classification of Educational Goals* (Vol. Handbook I, Cognitive Domain). London: Longman.

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

Chemers, M. M., Garcia, B. F., & Hu, L.-t. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology, 93*(1), 55-64.

Elias, S. M., & Loomis, R. J. (2000). Using an academic self-efficacy scale to address university major persistence. *Journal of College Student Development*(41), 4.

Holladay, C. L., & Quinones, M. A. (2003). Practice variability and transfer of training: The role of self-efficacy generality. *Journal of Applied Psychology, 88*(6), 1094-1103.

Klobas, J. E., Renzi, S., Francescato, D., & Renzi, P. (2002). Meta-response to online learning / Meta-risposte all'apprendimento online. *Ricerche di Psicologia, 25*(1), 239-259.

Maurer, T. J., & Pierce, H. R. (1998). A comparison of Likert scale and traditional measures of self-efficacy. *Journal of Applied Psychology, 83*(2), 324.329.

Owen, S. V., & Froman, R. D. (1988, April 6-8). *Development of a college academic self-efficacy scale.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleands, LA.

Pajares, F., Hartley, J., & Valiante, G. (2001). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and evaluation in counseling and development, 33*, 214-221.

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.

Smith, E. V., Wakely, M. B., De Kruif, R. E. L., & Swartz, C. W. (2003). Optimizing rating scales for self-efficacy (and other) research. *Educational and Psychological Measurement, 63*(3), 369-391.

Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science, 26*(1-2), 127-140.

University of Victoria Counselling Services Learning Skills Program. (2003). *Bloom's Taxonomy*. Retrieved 9 May, 2005, from www.coun.uvic.ca/learn/program/hndouts/bloom.html

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytica review. *Educational Psychologist, 39*(2), 111-133.

Wood, R. E., & Locke, E. A. (1987). The relation of self-efficacy and grade goals to academic performance. *Educational and psychological measurement, 47*, 1013-1024.

Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in Changing Societies* (pp. 202-231). Cambridge, UK: Cambridge University Press.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy belifs and personal goal setting. *American Educational Research Journal, 29*, 663-676.

Appendix 1: Study I activities

organizing	I am able to organize my activities so that I can meet [*most*, *all*] course requirements.
remembering[a]	Soon after the end of a lesson, I am able to remember [*most*, *all*] of the key concepts.
remembering[a]	After sitting an exam, I am able to remember [*most*, *all*] of the key concepts covered in the course.
understanding[a]	I can understand [*most*, *all*] of the concepts covered in my course.
understanding	I can [*usually*, *always*] interpret news reports related to a topic I am studying.
explaining[a]	I am able to explain to my fellow students, in a way they can understand [*some*, *many*, *all*] of the key concepts covered in a course.
connecting	When I find something new about a topic that I am studying, I am [*usually*, *always*] able to connect it with other things that I know about the topic.
updating	I [*usually*, *always*] know how to get up to date on a topic if my knowledge of it is dated

[a] Concepts derived from Wood and Locke (1987).

Appendix 2: Study II activities

	Activity	Theoretical basis[a]	SEL sub-scale and dimension[b]
1	I am able to organize my activities so that I can meet [*most*, *all*] course deadlines.		
2	Soon after the end of a lesson, I am able to remember [*most*, *all*] of the key concepts.	Info Processing	Info Processing: knowledge
3	I can understand [*most*, *all*] of the key concepts covered in my course.	Info Processing	
4	I am able to explain to my fellow students, in a way they can understand, [*most*, *all*] of the key concepts covered in a course.	Info Processing	
5	After sitting an exam, I am able to remember [*most*, *all*] of the key concepts covered in the course.	Info Processing	
6	When I find something new about a topic that I am studying, I am [*usually*, *always*] able to connect it with other things that I know about the topic	Info Processing	Info Processing: synthesis
7	When the media carries a report about something that I am studying, I can [*usually*, *always*] use my knowledge of the subject to interpret the report		
8	I [*usually*, *always*] know how to get up to date on a topic if my knowledge of it is dated	Info Processing	Info Processing: analysis
9	Even when I haven't participated in a lesson, I can [*usually*, *always*] understand the concepts covered in the lesson by reading a text book	Info Processing	
10	I am [*usually*, *always*] able to find material in the library about a subject that interests me	Finding	Finding: library
11	I am always able to find more detailed information on the Internet for a topic that [*interests me*, *does not interest me at all*].		
12	I am [*rarely*, *never*] embarrassed to ask the teacher for clarification		
13	I am [*usually*, *always*] able to identify the most appropriate person to help me resolve a problem related to my study		

.../

...Appendix 2 continued

	Activity	Theoretical basis[a]	SEL sub-scale and dimension[b]
14	I am [*usually*, *always*] able to evaluate the quality of fellow group members' contributions when I participate in group activities	Info Processing	
15	I am [*usually*, *always*] able to relate the notes I have made during a lesson with the topics covered in the course text or readings		
16	It is [*usually*, *always*] easy for me to understand new information, even on a topic that does not interest me very much	Info Processing	Info Processing: understanding
17	It is [*usually*, *always*] easy for me to connect new information about a topic that interests me with other pieces of information		
18	During a course, if we are given a new task to complete, I can [*usually*, *always*] complete it by applying the knowledge that I obtained from lessons		
19	Soon after the end of a lesson, I am [*usually*, *always*] able to distinguish the most important concepts from concepts of less importance	Info Processing	Info Processing: evaluation
20	If, as part of a course, I participate in a forum or online discussion, I am [*usually*, *always*] able to identify those message which will improve my understanding of the material covered in the course		
21	I am [*usually*, *always*] able to decide whether to go to the library or use the web, based on the type of information that I am seeking	Finding	Finding: distinguishing sources
22	I [*usually, always*] find it easy to join a group of fellow students to study or complete course activities	Finding	
23	I am [*usually*, *always*] able to identify useful information on the web for an essay	Finding	Finding: web
24	I am [*usually*, *always*] able to use the library and library services to select appropriate books and articles for an essay	Finding	Finding: writing
25	After a lesson, I am [*usually*, *always*] able to integrate concepts described by the teacher with those presented in course texts and readings	Info Processing	
26	When I write an essay for a course, I am [*usually*, *always*] able to incorporate knowledge gained from other sources	Info Processing	
27	I am [*usually*, *always*] able to help other students solve problems based on concepts described in a lesson	Info Processing	Info processing: application

[a] For items from the 18 item Rasch analysis solution. [b] For items included in 10 item SEL scale.