Exploration of Instruments Measuring Concepts of Graduateness in a Research University Context

J.M. Steur, E.P.W.A. Jansen *W.H.A. Hofman*

University Centre for Learning & Teaching, University of Groningen

The local Gaining Early Awareness and Readiness for Undergraduate Programs (GEARUP) partnership serves 11 K-8 schools with the lowest achievement scores and the highest poverty rates in a large Midwestern urban district. Recently, GEARUP launched a specially designed teaching program, Mathematics Enhancement Group (MEG), for underachievers in mathematics. To maximize the efficiency of classroom teaching, MEG, a pull-out program, was created to serve students with disorderly behavior but with a strong potential to improve their mathematics skills.

The pull-out program positively affected other students' achievement in the classroom, too. Overall proficiency level from the Ohio Achievement Test improved from 5.6 % of the previous year to 8.1%, while the district proficiency level decreased from 46.3 % to 38.7 %. The outcome of this research shows that MEG was effective for improving standardized test scores and study habits for the participants. Simultaneously, the setting appeared helpful in correcting students' classroom behavior.

Theoretical Background of Graduateness

Graduateness is considered an important generic learning outcome of university education (Barrie, 2004, 2005, 2006; Glover et al., 2002). However, this concept is difficult to measure as no theoretical model exists. In another article, Steur and colleagues (2010) proposed a theoretical model which, unlike other models for generic learning outcomes, explicitly separates graduateness from employability skills. As graduateness consists of concepts that are abstract and difficult to detect, including employability skills in the

definition leads to a greater focus on employability skills, which are more concrete and obvious. This paper focuses on the following research question: Can the instruments that are selected to measure concepts of graduateness be used in the context of a research university?

First, the theoretical framework of graduateness is briefly introduced. This serves as the background for the instrument selection. The different elements of graduateness are explained and the development towards graduateness is elucidated. Second, the instruments that measure the different aspects of graduateness are reviewed and selected. Some necessary adjustments to item formulations are made here. The exploration of these instruments in the context of a research university is then analyzed in the results section. This exploration is performed by means of factor analyses and reliability analyses. Finally, the results are discussed.

Introduction to Theoretical Framework

Reflective thinking is a key aspect when considering the literature on graduateness and higher cognitive learning outcomes. Reflective thinking is the ability to construct your own knowledge by collecting evidence in social interaction with others (King & Kitchener, 2004). Several scholars have stressed the importance of reflective thinking as an outcome of university education (Kember et al., 2000; King & Kitchener, e.g. 2004; Procee, 2001, 2006). However, it is our claim here that simply focusing on reflective thinking does not do justice to the complexity of graduateness. We have therefore explored the concept of graduateness to identify other important elements.

In a literature search another three concepts emerged related to graduateness: Scholarship, moral citizenship and lifelong learning. An important aspect here is the notion that graduateness is a specific outcome of academic intellectual development. Concretely, this means that graduateness only

occurs once a student has reached a specific stage in reflective thinking. It is argued here that this is accompanied by high achievement in at least one of the other elements of graduateness. Different developmental stages can be distinguished in these four domains. The separate concepts need to be operationalized before the theoretical model can be explored. They are then explored for their use in this specific context, as these instruments were not designed for use in research universities across disciplines. This is the central question in this article: To what extent can the selected instruments be used in a research university context?

Instrument Selection

This section reviews and selects the instruments for measuring the four elements of graduateness. The selection is made in the light of future application of the instruments. As the group of students wishing to enroll in Master's programmes becomes more diverse, pre-Master's programmes are being developed to prepare students from institutes other than research universities. An appreciation of where students stand in their development towards graduateness is vital for achieving a better fit between student needs and the pre-Master's programmes. Given the large numbers of students enrolling in these programmes, the best type of instrument is one that can be administered to large groups. There are other potential applications too, such as an opportunity to monitor student development concerning graduateness for, say, accreditation processes.

The literature search yielded a wide range of instruments to measure the four elements of the graduateness model. Given the number of instruments and the opportunity they provide for building on existing knowledge, we opted in this study to make a selection from these instruments rather than to develop entirely new ones. The selection procedure was governed by a number of criteria that the instruments

had to meet, bearing in mind the practical implications for the questionnaire. First, the instruments had to be administrable to large groups, resulting in minimum time investment for both respondents and researchers. Second, the instruments had to be applicable in a university context, or able to be adjusted for such a purpose. Third, as development is a component of the theoretical model, the instruments had to clearly show a student's development. The instrument selection is presented here, based on these criteria. In order to minimize the load on students and to ensure higher response rates, the instruments were critically assessed beforehand. If necessary, changes were made to items or items were deleted. This was necessary because of differences between the Dutch situation where the instrument would be used and the situation for which the instrument was developed.

Although King and Kitchener's (1981) Reflective Judgment Interview has been widely used in the field of reflective thinking, it had not been converted into a questionnaire. Kember and colleagues (2000) therefore developed a questionnaire – the Reflective Thinking Questionnaire – to measure this aspect. Its theoretical basis is derived from Mezirow (Kember et al., 2000), who distinguished different levels of reflective and non-reflective action. Kember and colleagues (2000) explicitly excluded affective components of reflective thinking because they felt that their instrument should focus on the level of reflective thinking actually attained. The instrument consists of 16 items distributed over 4 scales: habitual action, understanding, reflection and critical reflection, with reported reliabilities ranging from .58 to .91 (Kember & Leung, 2003; Phan, 2006). The scales suggest development towards reflective thinking. The four-factor structure of the Reflection Questionnaire was verified in a confirmatory factor analysis by Mann, Gordon and MacLeod (2009), who reported reliability estimates from α = .62 (habitual) to α = .76 (understanding). As all items

seem to fit the research university situation, no adjustments to the items were made.

Scholarship involves both the skills necessary to perform research and a scientific attitude. These aspects are commonly investigated in secondary education (e.g. Blalock et al., 2008; Gardner, 1995; Lederman et al., 2002; Noll, 1935; Stokking et al., 2004), but scholarship within universities has received less attention from researchers. Only a few studies have addressed student self-efficacy in performing specific research tasks. We found only four instruments, which focus on different elements of scholarship. One thing they have in common is that they feature a list of research-related activities. One of the instruments (Vodopivec et al., 2002) is specially designed to measure attitudes of first-year medical students. The content and context makes it unsuitable for general university-wide application.

Both the Self-Efficacy in Research Measure (Philips & Russell, 1994, as cited in Forester et al., (2004)) and the Research Attitudes Measure (O'Brien et al., 1998) fail to make development explicit. They distinguish respectively four and six dimensions, and development is expected to appear in all dimensions, rather than the dimensions reflecting a specific level of development. This complicates the development measures. Nor do the instruments make a clear distinction between scholarship and employability. The Self-Efficacy in Research Measure in particular builds heavily on the more specific employability. The Research Self-Efficacy Scale/RSES (Bieschke et al., 1993) seems to have the closest connection to scholarship as understood in the current model. It also has the potential to make development clearly evident. In a validation study of instruments for research self-efficacy, Forester and colleagues (2004) examined the structure of the RSES. Unfortunately, they were unable to confirm the intended structure in their study. They do

suggest, however, treating the instrument as a single-factor scale.

The instrument was developed for use in the social sciences. Some items were too representative of the social sciences to be appropriate for other disciplines (e.g. those items with a strong emphasis on statistical data analyses). Where possible, these items were rewritten in more general terms, such as a general way of dealing with the data to answer a particular research question. If rewriting was impossible or would lead to statements that were already included, the items were omitted. The instrument was developed in the early 1990s, when computers were used less frequently than today. The statements on computer skills were rather general and were therefore not expected to have any discriminative value (e.g. 'use computer software to prepare texts, like word processing'). These items were also omitted. An item on contributing to the body of knowledge within the discipline was added. The modified version of the instrument contains 46 items.

The idea of *moral citizenship* refers to two aspects of life: Moral reasoning ability and participating in and contributing to society. A substantial number of instruments have been developed in this field, some of which are also suitable for administration to adults. The instruments can be divided into those that are grounded in Kolhberg's moral theory and those that are not. Kohlberg (1973) distinguishes different stages in moral development at which an individual makes moral judgments. These judgments become increasingly sophisticated as an individual develops. In earlier stages moral judgments are quite dualistic while at higher stages an individual takes different points of view and pieces of evidence into account before judging a situation. The instruments that are developed in the Kohlbergian (1973) tradition have a similar structure. A dilemma is presented in a short story, and the respondent is asked to state how (s)he

believes the main character in the story should act. This structure differs substantially from the regular questionnaire format. We therefore decided not to use this type of instrument. Although commonly used in moral reasoning studies, the Defining Issues Test (Rest et al., 2000; Rest et al., 1999; Rest, 1975) requires students to formulate their own answers, which complicates data processing for large groups and makes it less appropriate for use in the present study.

The other instrument in this tradition has its own complication. The Sociomoral Reflection Objective Measure (Gibbs et al., 2007; Gibbs et al., 1984; Gibbs et al., 1982) does not cover the entire developmental range. Two instruments seem potentially suitable for measuring development to a limited extent. Like the instruments in the Kohlbergian tradition, the Moral Authority Scale (see Bringle et al., 2004 for references) consists of six moral issues in a story format. In the Visions of Morality Scale/VMS (Shelton & McAdams, 1990) the situations are outlined in short sentences, which better fits the common item structure of a questionnaire. The items are distributed over three dimensions: private, interpersonal and social. It seems possible to significantly reduce the number of items in this instrument, which is not possible in the story-type instruments. After critically reviewing the instrument, we found 36 situations that were applicable in a university setting. Some needed minor modifications.

Although the VMS consists of three subscales, Punzo (1993, as cited in Fry (2005)) and Fry (2002, as cited in Fry, 2005) argued that one subscale, the social morality scale, should be omitted. These two studies deal with the remaining scales as a single scale with sufficient reliabilities (α = .86 and α = .88 respectively). An alternative suggestion was to use all items as one scale (see Bringle et al, 2004). This would be preferable for the present study, since moral citizenship is believed to cover all three dimensions.

The idea behind lifelong learning is that, at a certain stage in their development, students are capable of directing their own learning (among others, Paris & Newman, 1990). This fits more within the Self-regulated Learning Approach tradition (e.g. Pintrich, 2004; Boekaerts & Corno, 2005) than the Student Approaches to Studying tradition (Biggs, 1999; Marton & Saljö, 2005). Self-regulation is not an explicit part of the latter tradition, however it is an important element of graduateness. The instruments in the Self-regulated Learning Approach tradition therefore seem more appropriate for the graduateness model as used here. Although these instruments are not explicitly designed to measure development, the subscales suggest this possibility.

The Motivated Strategies for Learning Questionnaire/MSLQ (Pintrich et al., 1991) is commonly used in self-regulation research. It consists of a number of subscales on motivation and learning strategies. We used only five subscales: rehearsal, elaboration, organization, critical thinking and metacognitive self-regulation, which cover 31 items. The instrument has been validated in several countries and situations. The MSLQ subscales used are all cognitive elements of self-regulation, whereas the theoretical foundation of the MSLQ explicitly expects integration of both cognitive and affective elements.

Two scales concerning goal orientation (Duda, 1992; Van Yperen & Diderich, 1998) were added to satisfy this. These are mastery orientation and performance orientation, which are covered by a total of 18 items. The option of limiting the number of items was the main reason for choosing this instrument over the MSLQ motivation subscales. Only minor modifications were made to both instruments. In addition, this instrument has a less complex structure and it has already been validated in a Dutch situation by Van Yperen and Diderich (1998).

Method

All instruments were translated into Dutch and then back into English to check the appropriateness of the Dutch translation. As described in the previous section, some adjustments were made to ensure a better fit between the items and the context of a research university and to prevent item bias (Van der Vijver & Hambleton, 1996). Some items were omitted, as their specific character made it impossible to rewrite them for a generic university context. In the domain of scholarship, for example, the instrument was oriented to the social sciences. Items that dealt with data analyses in the social sciences were omitted.

At a Dutch research university, students from two different disciplines – social sciences and the humanities – were invited to participate in this study. Unfortunately, it was not possible to contact the students from each discipline in the same way. The social science students were invited by e-mail to fill in the online questionnaire. If they did not respond, they received at most two reminders. The humanities students were asked to participate through a notice placed in their electronic learning environment. It was not possible to send these students a reminder. This resulted in a different response rate for the two groups.

The length of the questionnaire appeared to affect the response rate. Some students did not complete the entire questionnaire. The students who completed most of it (e.g. they left out no more than 10 answers) were selected for the analyses. Most of the questionnaire was completed by 108 out of 420 social sciences students and by 43 out of 728 humanities students. However, we know that only 88 humanities students had read the invitation to participate, as they had viewed at least the opening page of the online questionnaire. For each analysis the missing values were deleted pairwise.

Results

The aim of this article is exploring the usability of the instruments in a research university context. To what extent are the instruments applicable within a research university? Factor analyses with varimax rotations were performed to explore the intended structures, together with a calculation of reliability estimates in the form of Cronbach's alphas. The analyses were performed for each instrument separately. For the factor analyses, the number of factors were pre-set within the analyses, according to the intended number of factors. In the tables below, the grey cells indicate the intended factor that these items should load on. A brief summary is provided of the items, but these are not the exact phrases used in the questionnaire. Summaries were preferred to the exact wording for reasons of readability and because of constraints concerning table layout.

The factor analysis on the reflective thinking instrument neatly reflects the intended four-factor structure. The reliability measure indicates sufficient internal consistency within the scales, with the exception of the critical thinking scale, which has a relatively low Cronbach's alpha. Table 1 shows the factor loadings on the four factors and the reliability estimates.

Table 2 reports the factor loadings for the four-factor solution of the Research Self-Efficacy Scale, together with the reliability estimates. The factor that the items are intended to load on is shaded in grey. The intended structure was not found in an explorative factor analysis, which is in line with the findings of Forester and colleagues (2004) concerning the structure of this instrument. They found that the items loaded best on a single factor structure. This solution was also considered here, even though the reliabilities of the intended scales are good, with the exception of the Early Tasks scale. It appears that almost all items have high loadings on the single factor. The entire scale has a reliability estimate of .95, partly

due to the large number of items

The Visions of Morality Scale assumed a three-factor structure but this was not found in the explorative factor analysis (see Table 3). Furthermore, Cronbach's alphas – ranging from .220 - .446 – for the intended scales suggested no internal consistency within the scales. Following Bringle and colleagues' (2004) suggestion to treat all items as a single factor does not lead to a satisfactory solution. Nor does omitting the social scale items and allowing a two-factor solution (not reported here).

The factor analyses on the MSLQ did not reveal the intended five-factor structure. Items were deleted from the MSLQ subscales Elaboration and Metacognitive Self-Efficacy to improve reliability. Selection was performed on both item content and statistical measurements (improvement in Cronbach's alpha when items were omitted). No hierarchy was found, which is in line with Pintrich's theory on Self-Regulated Learning. The reliabilities are presented in Table 4.

Table 1
Factor loadings and Cronbach's alphas for Reflective Thinking Questionnaire

#Item	Habitual	Under-standing	Critical Thinking	Reflective Thinking	Reliability *(intended scales)*
Doing without thinking	.765				.770
Repeating many times	.742				
Remembering handouts	.786				
Following what lecturer says	.770				
Understanding concepts		.777			.814
Understanding content		.630			
Understanding material		.695			
Thinking about material		.648			
Questioning the way others do things			.518		.601
Thinking of alternative ways			.724		
Reflecting to improve			.659		
Reappraising experience			.676		
Changed way of looking at myself				.649	.777
Strongly held ideas are challenged				.777	
Changed normal way of doing things				.773	
Discovered faults in beliefs				.817	
Eigenvalue	2.002	2.788	1.989	2.565	
% variance explained	13.35	18.59	13.26	17.10	

Only factor loadings over .40 are presented.

Table 2

Factor loadings and Cronbach's alphas for Research Self-Efficacy Scale

# Item	Concept-ualization	Implement ations	Early tasks	Present-ing results	Reliability (*intended scale*)	Single factor
Assessing quality of scientific articles	.591			.414	.871	.615
Participating in brainstorming session		.583				.493
Discussing ideas with students				.458		
Consulting senior researchers		.556				.547
Knowing when to stop looking for lit.	.668					.494
Knowing when to stop developing ideas	.547					.525
Relating various ideas and findings	.620					.653
Identifying useful research areas	.461		.432			.606
Putting forward logical reasoning	.400		.439			.667
Drawing up research questions	.544					.628
Drawing up research proposal		.459				.519
Editing texts to make them concise						.448
Presenting ideas to study group				.707		.490
Revising research ideas based on criticism			.492			.580
Selecting appropriate research design	.702					.617
Being flexible in developing alternative research strategies	.465		.437			.721
Working independently in research group		.501			.881	.430
Selecting best way to collect data	.721					.675
Selecting appropriate analytical techniques	.666					.736
Obtaining permission		.709				.545
Ensuring access to required objects		.416				.618
Instructing assistants		.784				.531

Observing research procedures			.615			.607
Ensuring reliable data collection			.577			.631
Supervising assistants		.780				.432
Awareness of appropriateness of data coll.	.407					.581
Preparing data for analyses	.634					.664
Processing results into tables, figures etc			.445			.572
Using computer for data analyses			.403			.455
Drawing reliable conclusions from results	.640					.680
Presenting results visually				.527		
Obeying ethical research principles						.644
Knowing where to start in literature	.733		.505		.530	.552
Finding references						.581
Finding articles in other libraries	.545					.458
Producing professional research report	.407				.781	.626
Discussing findings in light of current literature			.404			.473
Identifying and reporting limitations			.753			.607
Describing implications			.733			.612
Presenting results to research group				.756		
Presenting results at a conference		.574				
Defending results to critical audience		.498		.503		.457
Writing publication for academic journal		.482				
Finding and applying for grant						
Keeping references well organized						.652
Contributing to development of discipline						.532
Eigenvalue	6.994	5.711	5.930	3.491		14.122
% variance explained	15.20	12.42	12.89	7.588		30.70

Only factor loadings over .40 are presented.

Table 3
Factor loadings and Cronbach's alphas for Visions of Morality Scale

# Item	Private	Interper-sonal	Social	Reliability (*intended scale*)	Single factor
Giving euro you found to charity			.549	.224	.507
Smiling to make others feel better					
Helping motorist stuck in snow					
Taking over colleague's shift	.431				
Taking clothes to Salvation Army					
Donating money from lottery to street party					
Donating money to family after fire	.653				
Change money for parking	.573				
Giving blood when asked to on the street					
Giving to charity					
Filling in questionnaire		.606			
Volunteering at annual open day					
Helping at evening with celebrity speaker				.446	
Watching basketball match	.502				
Offering concert ticket to exchange student			.404		.437
Helping fellow student with difficult course		.598			
Giving money to student who lost wallet					
Giving blood to student after an accident			.513		
Being friendly to lecturer whose father is ill					.471
Helping neighbour with shopping		.454			
Helping organize high school reunion			.542		
Apologizing after heated discussion		.442			
Helping blind lady on the street					.479

Offering lift to job fair if petrol is paid for				.442
Helping lady to carry heavy bags				
Participating in sponsored run for charity			.220	
Voting for tax rise to benefit the poor			.551	.425
Going to shop that sells sustainably produced goods			.470	-.423
Not going to a film that encourages sexist and violent behaviour				
Helping prepare meal for homeless			.429	
Writing protest letters	.511			.460
Eating less and donating money to world food programme			.562	.633
Not eating in restaurant that bars ethnic minorities			.509	.412
Telling student with money problems about a vacancy at work				
No longer buying favourite snack				
Introducing more sustainable lifestyle in dorm			.621	.550
			.490	.435
Eigenvalue	2.875	2.591	3.448	4.405
% variance explained	8.22	7.40	9.85	12.24

Only factor loadings over .40 are presented

Table 4

Factor loadings and Cronbach's alphas for Motivated Strategies for Learning Questionnaire

# Item	Rehearsal	Elaboration	Organization	Critical thinking	Metacognitive Self-Regulation	Reliability *(intended)*
Practising by saying the material to myself over and over		.472				.671
Reading class notes and course readings over and over	.405					
Memorizing keywords						
Making and memorizing lists						
Pulling together information from different sources		.626				.603
Relating ideas in this subject to other courses		.420		.494		
Relating material to what is already known		.478		.474		
Writing brief summaries			.751			
Understanding by making connections between lectures and reading material		.749				
Applying ideas from course reading to other class activities		.449				
Making outline of material to organize thoughts			.781			.730
Going through readings and class notes to find most important ideas		.523				
Making charts, diagrams, etc. to organize course material			.838			
Making outline of important concepts			.795			
Questioning things in the course and deciding if they are convincing		-.422		.541		.783
Deciding if there is good support for theories,				.556		
				.705		

interpretations etc.				.794		
Using course material as starting point to develop own ideas				.507		
Playing around with own ideas relating to course content						
Thinking of alternatives for assertions or conclusions						
Often missing important things (reversed)						.601
Making up question to focus reading	.615					
In the event of confusion, going back and trying to figure it out	-.593	.592				
For difficult texts, changing the way you read						
Skimming new reading material	.740					
Asking questions to make sure you understand					.693	
Changing way of study to fit course material and teaching style						
Reading course material without knowing what it is all about (reversed)		.680			.690	
Thinking about and deciding what you are supposed to learn		.553				
Determining which elements are not understood well		.537				
Setting goals to direct study activities						
Confusion with note-taking during lecture is sorted out later						
Eigenvalue	2.359	3.734	3.559	3.538	1.682	
% variance explained	7.61	12.04	11.48	11.41	5.43	

The factor analysis for the Task and Win Orientation in Sports Questionnaire neatly revealed the intended two-factor structure. This is presented in Table 5. The Cronbach's alphas indicate good internal consistency for the scales.

Table 5
Factor loadings and Cronbach's alphas for Task and Win Orientation in Sports Questionnaire

Item	Mastery	Performance	Reliability
Improving yourself in some respects	.649		.873
Observing self-improvement	.720		
Attaining new knowledge or skills through hard work	.625		
Learning something new that is great fun	.643		
Attaining something that inspires you to do even better	.766		
Doing your very best	.727		
Performing to your best ability	.702		
Acquiring new knowledge or skills that used to be troublesome	.582		
Learning something that inspires you to go on			
Performing to your maximum ability			
Perfectly mastering new knowledge or skills			
Others do not perform at your level		.860	.903
You perform better than others		.825	
You can demonstrate clearly that you perform better than others		.818	
You have contributed the most to a group assignment		.609	
You are the best		.828	
Others fail, but you don't		.572	
You succeed where others do not		.817	
You are the only one to master certain knowledge or skills		.777	
Eigenvalue	4.913	5.316	
% explained variance	23.40	25.32	

Only factor loadings over .40 are shown.

Conclusion and Discussion

The purpose of this study was to explore the use of instruments selected to operationalize the theoretical model on graduateness in the context of a research university. The instruments shown to be applicable within Dutch research universities will be used in a follow-up study to test the theoretical model empirically. First, the instruments were selected based on their appropriateness for the theoretical model and their suitability for large-group administration. This latter criterion was necessary due to the instrument's practical implications. Second, empirical data were presented on the usability of the instruments in the context of a Dutch university. This section summarizes the results and draws conclusions on the suitability of the different instruments in this particular situation.

Five instruments were selected to measure four concepts. The selected instruments were adjusted, where necessary, to match the context of a Dutch university. This modified version was then administered to students of the Social Sciences and the Humanities across all years (from first-year Bachelor's students to Master's students). Factor analyses and reliability analyses were used in this explorative study.

For the Reflective Thinking Questionnaire the intended structure was found, with reasonable reliability estimates. This confirms the usability of this instrument in a Dutch university context. Because this instrument is already as compact as possible, no recommendations are made to reduce the number of items.

The four-factor structure of the Research Self-Efficacy Scale could not be reconstructed with the current data. However, this was also found in another replication study in the US. The alternatively proposed single factor solution did lead to satisfactory results. Almost all items loaded on that factor and the reliability estimate was high. For

further use of this instrument it is suggested that the number of items be reduced. One important drawback of using a single factor to measure scholarship is that it is more difficult to make the development component apparent, which is a key aspect of the theoretical model for graduateness. This problem could be solved by looking for a hierarchy in the items by means of item-response analysis. This could be considered when selecting items for the shortened version of the instrument.

The use of the Visions of Morality scale could be problematic. The intended structure could not be replicated, nor could a single factor solution be produced covering all items. Considering the low reliability estimates, it is not advisable to use this instrument in its present form. If possible, considerable adjustments need to be made. We should also ask ourselves whether looking at moral behaviour in everyday situations is the best guide to moral academic behaviour. However, given the available instruments, the VMS was the best choice (although suboptimal). We should also consider whether other additional instruments are needed to measure this complex concept. Another option is to step away from the intended subscales and freely explore the items for underlying constructs.

Two instruments were selected for Lifelong Learning, one to measure the cognitive aspects and one to measure the affective aspects. The cognitive aspects were measured using five subscales of the Motivated Strategies for Learning Questionnaire. The intended structure could not be replicated. However, the reliability estimates were quite good. Relatively low reliabilities for some subscales were found elsewhere (Artino, 2005), with the suggestion that lower reliabilities were preferred to an increase in the number of items. We recommend checking whether the elimination of a couple of items would improve these reliability estimates. Pintrich (2000) suggests that the subscales are strategies that

everyone adopts, regardless of their learning routines. This could explain why it is hard to find a rotation with maximum differences between the five factors. They are somehow related to each other. The affective aspects were measured using an adjusted version of the Task and Win Orientation in Sports Questionnaire. This structure could be replicated easily in the available data. The reliability estimates were also good. This instrument could therefore easily be used. However, given the number of instruments to be administered for measuring graduateness, we recommend taking a close look at which items could be omitted.

An important limitation of this study is that it is confined to two distinct disciplines within a research university. The sample was selected based on availability and the general character of the selected disciplines. The variety in responses was obtained by considering all year groups within the discipline. The usability of the instruments in the research university context was confirmed for all instruments, with the exception of the Visions of Morality Scale. We suggest a free exploration to detect any underlying constructs within the items. For the other instruments, reducing the number of items should be considered because the long version led to low response rates.

References

Artino, A.R. (2005) Review of the motivated strategies for learning questionnaire. Retrieved from: http://www.eric.ed.gov [ED499083]

Barrie, S.C. (2004). A research-based approach to generic graduate attributes policy. *Higher Education Research and Development*, 23 (3), 261 – 275.

Barrie, S.C. (2005). Rethinking generic graduate attributes. *HERDSA News* 27 (1), 1-6.

Barrie, S.C. (2006). Understanding what we mean by the generic attributes of graduates. *Higher Education, 5* (2), 215-241.

Bieschke, K.J. (1993). Research self-efficacy beliefs and research outcome expectations: Implications for developing scientifically minded psychologists. *Journal of Career Development, 14* (1), 77-91.

Biggs, J. (1999). What the student does: teaching for enhanced learning. *Higher Education Research & Development, 18* (1), 57-75.

Blalock,,C.L., Lichtenstein, M.J., Owen, S., Pruski, L., Marshall, C. & Toepperwein, M. (2008). In pursuit of validity: A comprehensive review of science attitude instruments 1935-2005. *International Journal of Science Education, 30* (7), 961-977.

Boekaerts, M. & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54* (2), 199-231.

Bringle, R.G., Philips, M.A. and Hudson, M. (2004). *The measure of service learning: Research scales to assess student experiences.* Washington, DC: American Psychological Association.

Duda, J.L. (1992). Motivation in sport settings: A goal perspective approach. In Roberts, G.C. (Ed.) *Motivation in sport and exercise* (pp.57-91) Champaign, IL: Human Kinetic Books.

Forester, M., Kahn, J.H. & Hesson-McInnis, M.S. (2004). Factor structures of three measures of research self-efficacy. *Journal of Career Assessment, 12* (1), 3-16.

Fry, H.L. (2005). *The application of virtue ethics to the practice of counselling Psychology.* Doctoral Dissertation at Ohio State University.

Gardner, P.L. (1995). Measuring attitudes to science: Unidimensionality and internal consistency revisited. *Research in Science Education 25* (3), 283-289.

Gibbs, J.C., Widaman, K.F. & Colby, A. (1982) Construction and validation of a simplified, group-administerable equivalent to the moral judgment interview. *Child Development, 53* (4), 895-910.

Gibbs, J.C., Arnold, K.D., Morgan, R.L., Schwartz, E.S., Gavaghan, M.P. & Tappan, M.B.. (1984) Construction and validation of a multiple-choice measure of moral reasoning. *Child Development, 55* (2), 527-536.

Gibbs, J.C., Basinger, K.S., Grime, R.L. & Snarey, J.R. (2007) Moral judgment development across cultures: Revisiting Kohlberg's universality claims. *Developmental Review, 27* (2), 443-500.

Glover, D., Law, S. & Youngman, A. (2002). Graduateness and employability: student perceptions of the personal outcomes of university education. *Research in Post-Compulsory Education, 7* (3), 293-306.

Kember, D., Leung, D., Jones, A., Loke, A., McKay, J., Sinclair, K., Tse, H., Webb, C., Wong, F., Wong, M., & Yeung, E. (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment and Evaluation in Higher Education, 25*, 381–395.

King, P.M. & Kitchener, K.S. (1981). Reflective judgment: Concepts of justification and their relationship to age and education. *Journal of Applied Developmental Psychology, 2* (2), 89-116.

King, P.M. & Kitchener, K.S. (2004). Reflective judgment: theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist, 39* (1), 5-18.

Kohlberg, *L.(1973).* The claim to moral adequacy of a highest stage of moral judgment. *Journal of Philosophy, 70,* 630–646

Lederman, N.G., bd-El-Khalick, F., Bell, R.L. & Schwartz. R.S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Teaching Science, 39* (6), 497-521.

Leung, D.I.Y. & Kember, D. (2003). The relationship between approaches to learning and reflection upon practice. *Educational Psychology, 23* (1), 61-71.

Mann, K., Gordon, J. & McLeod, A. (2009). Reflection and reflective practice in health professions education: A systematic review. *Advances in Health Sciences Education, 14* (4), 595-621.

Marton, F. & Saljo, R. (2005). Approaches to learning. In: Marton, F., Hounsell, D. & Entwistle, N., (Eds.) *The experience of learning: Implications for teaching and studying in higher education.* 3rd (Internet) edition (pp. 39-58). Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment.

Mokken R.J. (1971). *A theory and procedure of scale analysis.* Berlin: De Gruyter.

Noll, V.H. (1935). Measuring the scientific attitude. *Journal of Abnormal and Social Psychology, 30* (2), 145-154.

O'Brien, K. M., Malone, M. E., Schmidt, C. K., & Lucas, M. S. (1998, August). *Research self efficacy: Improvements in instrumentation.* Poster session presented at the annual conference of the American Psychological Association, San Francisco.

Paris, S.G. & Newman, R.S. (1990). Developmental aspects

of self-regulated learning. *Educational Psychologist, 25* (1), 87-102.

Phan, H.P. (2006). Predicting change in epistemological beliefs, reflective thinking and learning styles: A longitudinal study. *British Journal of Educational Psychology, 78*, 75-93.

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).* Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.

Pintrich, P.R. (2000). The role of goal orientation in self-regulated learning. In: Boekaerts, M., Pintrich, P.R. and Zeidner, M. (Eds.), *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic Press.

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

Procee, H. (2001). Competenties en onderwijs: een conceptuele analyse. [Competences and education: A conceptual analysis]. *Tijdschrift voor Hoger Onderwijs, 19* (4), 242-253.

Procee, H. (2006). Reflection in education: A Kantian epistemology. *Educational Theory, 56* (3), 237-253.

Rest, J.R. (1975). Longitudinal study of Defining Issues Test of moral judgment - Strategy for analyzing developmental change. *Developmental Psychology, 11* (6), 738-748.

Rest, J., Narvaez, D., Thoma, S.J. & Bebau, M.J. (1999). DIT2: Devising and testing a revised instrument of moral judgment. *Journal of Educational Psychology, 91*, 644-659.

Rest, J.R., Narvaez, D., Thoma, S.J. & Bebeau, M.J. (2000). A neo-Kohlbergian approach to morality research.

Journal of Moral Education, 29 (4), 381-395.

Shelton, C. M., & McAdams, D. P. (1990). In search of an everyday morality: The development of a measure. *Adolescence, 25*, 923-943.

Steur, J.M., Jansen, E.P.W.A & Hofman, W.H.A. (2010). *Graduateness: A theoretical framework.* Manuscript submitted for publication.

Stokking, K., van der Schaaf, M., Jaspers, J. & Erkens, G. (2004). Teachers' assessment of students' research skills. *British Educational Research Journal, 30* (1), 93-116.

Van Yperen, N. W. & Diderich, M.C. (1998). De knikkers of het spel? Verschillen tussen werknemers in doelorientatie, attributies van succes en motivatie. [Is it wining that counts or taking part? Differences in goal orientation, attributes of success and motivation for employees.] *Nederlands Tijdschrift voor de Psychologie, 53*, 76-84.

Vijver, F. van der & Hambleton, R.K.. (1996). Translating tests: Some practical guidelines. *European Psychologist, 2* (2), 89-99.

Vodopivec, I., Vulaklija, A., Hrabak M., Kresimir Lukic, I., Marusic, A. & Marusic , M. (2002). Knowledge about and attitude towards science of first year medical students. *Croatian Medical Journal, 43* (1), 56-62.