Higher Education Learning Outcomes: Assessment & Measurement Challenges

CHAIR: Gavin T L Brown, University of Auckland, New Zealand

Paper 1: Olga Zlatkin-Troitschanskaia, Johannes Gutenberg University Mainz,
Hans Anand Pant, Humboldt-Universität zu Berlin,
Miriam Toepper, Johannes Gutenberg University Mainz
Dimitri Molerov & Corinna Lautenbach, , Humboldt-Universität zu Berlin, Germany

Paper 2: Eduardo C. Cascallar, University of Leuven & Assessment Group International
Mariel F. Musso, University of Leuven & National Research Council (Argentina)
Eva Kyndt, University of Leuven

Paper 3: Gavin T L Brown & Makayla P. Grays, University of Auckland

Discussant: David Boud, Deakin University, Australia

Abstract: This symposium addresses the measurement and evaluation of higher education learning outcomes (HELO). There is global interest in the quality of student outcomes and a growing demand for evidence that undergraduate education provides value beyond simply being a means of career preparation and placement. HELOs consist of a range of key competencies that include disciplinary specific skills and knowledge, generic cognitive and communicative skills, and certain personal dispositions or attributes. These competencies are generally accepted as valued goals of undergraduate education and universities commonly argue that students acquire them as part of completing an undergraduate degree. However, while some argue that these competencies cannot and even should not be assessed, it is increasingly likely that evidence concerning these benefits will need to be provided. In addition to demonstrating accountability, universities themselves may use such evidence to monitor their own effectiveness. This symposium focuses on methods of assessing HELO by providing three distinct approaches. Paper 1 describes a national program of competence definition and modelling, Paper 2 describes three studies using neural network analysis to classify students according to academic performance, and Paper 3 reports a single-site repeated measure analysis of self-reported dispositions according to degree completion status. These three papers show the wide diversity of approaches being undertaken internationally. While each paper will report specific substantive results about factors contributing to various HELO, the focus of the symposium is to advance our understanding of methods of data collection and analysis concerning a wide range of HELO.

There is an intensifying interest in the quality of student experiences in higher education and a growing demand for evidence that undergraduate education provides additional value to alternative means of career education. Whether as ‘client’ or an ‘intellectual partner’, the student is expected to benefit from a university education in ways that meet society’s, as well as their own, expectations. The fundamental premise is that, over and above any career entry or economic benefits accruing to a person with a university degree (especially in selective professions), university education is intended to have certain impacts on an individual that could not be obtained in another way. These benefits consist of three major types: (a) in-depth knowledge and mastery of the content and methods of an established discipline, (b) advanced critical intellectual powers such as logic, reasoning, analysis, synthesis, problem solving and so on, and (c) enhanced personal attributes and dispositions favourably associated with a broader, more tolerant, engaged, and insightful perspective on humanity,

organisations, values, ideas, and beliefs. The expectation is that the university can demonstrate to funding and sponsoring agencies as well as society as whole, that value for money is obtained in the development of the kind of people society requires, not just in terms of technical skills, but also in terms of the qualities that they bring to social life and citizenship.

Arguments have been made that these key competencies cannot and should not be assessed (Harris, 2001). However, it is increasingly likely that such benefits may need to be made transparent to economic, political, and citizen sponsors of the university, and so, in some form or another, be assessable. Notwithstanding the importance of demonstrating accountability, the university itself needs to be able to monitor, maintain, and improve the quality of its educational impact on students. Considerable research has commenced into the evaluation of student cognitive and communicative competencies, although less is known about student acquisition of personal dispositions (Zlatkin-Troitschanskaia, Shavelson, & Kuhn, 2015). For example, extant systems for determining value added measures include: (a) the *Assessing Higher Education Learning Outcomes* (AHELO) program from the OECD, (b) the *Collegiate Learning Assessment* (CLA), (c) the *Student Experience in the Research University Survey* (SERU-S), (d) the *National Survey of Student Engagement* (NSSE), (e) the *Course Experience Questionnaire*, and (f) the *College Student Experience Questionnaire*. The relevance of these systems to the HELO curriculum outcomes of any specific university or program need to be evaluated. While extant systems may allow relatively rapid data collection, they may not have significant face validity. In response, James Madison University has designed and validated a set key competencies measures which it administers in an annual Assessment Day in which all students are required to participate (Zilberberg, Brown, Harmes, & Anderson, 2009).

Given the diversity of learning outcomes, sites, methods of data collection and analysis, it is opportune to bring together in this symposium three different threads of research into the measurement of HELO. Paper 1 overviews the first 5-year phase of the German research program Modeling and Measuring Competencies in Higher Education (KoKoHs). KoKoHs projects involved 24 alliances among researchers at almost 50 higher education institutions who conducted 70 single projects to model more than 40 domain-specific and generic competencies acquired at an academic level and transformed them into measurement models and instruments that were later tested empirically and validated. Paper 2 overviews 3 studies in which artificial neural network analysis was used to accurately predict classification of students into high, middle, or low academic achievement. These analyses identified that the academic level of a student depended on different predictor variables and suggests areas in which potential interventions could improve performance. Paper 3 is a single site, quasi-experimental repeated measures study of self-reported dispositions in which confirmatory factor analysis and invariance testing established robust models which showed substantive and statistically significant differences depending on degree attainment status.

Together these three papers provide interesting insights into robust approaches to the measurement and detection of HELOs. The studies show that certain kinds of HELOs are reliably associated with university degree acquisition and consequently provide potentially useful feedback to the design of university curricula to ensure students acquire intended outcomes. Furthermore, the measures of data collection and analysis have potential for generating evidence to external stake-holders that value to society is associated with degree completion in universities. The papers also demonstrate that detection and evaluation of HELOs is possible. It is hoped that this symposium will trigger important discussion as to appropriate directions in further evaluation and assessment of HELOs.

References

Harris, B. (2001). Are all key competencies measurable? An education perspective. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies* (pp. 222-227). Seattle, WA: Hogrefe & Huber.

Zilberberg, A., Brown, A. R., Harmes, J. C., & Anderson, R. D. (2009). How can we increase student motivation during low-stakes testing? Understanding the student perspective. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 255-278). Charlotte, NC: Information Age Publishing.

Zlatkin-Troitschanskaia, O., Shavelson, R. J., & Kuhn, C. (2015). The international state of research on measurement of competency in higher education. *Studies in Higher Education, 40*(3), 393-411. doi: 10.1080/03075079.2015.1004241

Valid Assessment of Higher Education Learning Outcomes in Germany

Olga Zlatkin-Troitschanskaia,
Johannes Gutenberg University Mainz
Hans Anand Pant
Humboldt-Universität zu Berlin
Miriam Toepper
Johannes Gutenberg University Mainz
Dimitri Molerov & Corinna Lautenbach,
Humboldt-Universität zu Berlin

Abstract: Over the past decade, there has been growing interest in various issues related to the provision of higher education. Policy-driven outcome-oriented reform strategies (such as the Bologna reform) have changed higher education in the long term. Growing internationalization of higher education as well as increasing global mobility of students call for greater transparency of, and valid information on, students' knowledge and skills. Various theoretical and methodological challenges in assessment arise from the immense diversity of degree courses, study programs, and institutions. Empirical findings on the effectiveness of higher education programs can serve as a basis for sustainable development and reforms at structural, organizational, and individual levels. A recent review of the literature has revealed a substantial lack of research on assessment practices in higher education, especially on domain-specific and generic competency models, measurement models, and valid methods to assess competencies. This state of international research in the area of higher education influenced the conceptualization and implementation of the German research program Modeling and Measuring Competencies in Higher Education (KoKoHs) (2011–2015). In KoKoHs, models of cognitive abilities and skills have been operationalized through measuring instruments and tested in empirical assessments. Results from the first five-year phase of KoKoHs indicate that the program generated evidence of the quality of developed models and instruments and of the reliability of information on the assessed competency constructs. In this presentation, we describe the aim and conceptual and methodological framework of KoKoHs and present the main results of the first phase of the program.

Introduction

Over the past decade, policy-driven outcome-oriented reform strategies (e.g., the Bologna reform, the Assessment of Higher Education Learning Outcomes feasibility study (AHELO) by the Organisation of Economic Cooperation and Development (OECD), and the European Association for Quality Assurance in Higher Education (ENQA)) have changed higher education in the long term, particularly in OECD countries. These changes can be attributed in part to the immense increase in access to higher education and to the effects of internationalization of study programs and mobility of students. These changes have led to an urgent need for international benchmarking standards to provide evidence of student learning outcomes in higher education that can be compared across institutions and countries (see Coates, 2014; Land & Gordon, 2013; Tremblay et al., 2012; Liu, 2011).

In the course of internationalization, stakeholders (e.g., policy makers) seek information on accountability and efficiency of higher education institutions and the quality of their programs and courses. They also seek evidence of positive student learning outcomes. To provide this information, research on higher education requires a sound theoretical and empirical basis. Effectiveness of higher education programs and courses can be determined only through formative and summative assessment of learning outcomes and important influence factors such as basic generic and domain-specific skills and competencies. Findings from empirical research on the effectiveness of higher education programs can serve as a basis for sustainable development and reforms at structural, organizational, and individual levels. Higher education is still underrepresented in international empirical research,

and the related literature is relatively limited compared, for example, to the literature on teaching and learning (Land & Gordon, 2013).

Background of the KoKoHs research program

A comprehensive and systematic review of the state of international research in the field of learning outcomes assessment in higher education (Kuhn & Zlatkin-Troitschanskaia, 2011) revealed that some approaches had been taken to stabilizing the structure of empirical research on higher education, but this kind of research was still largely underrepresented. This finding led to the conceptualization and implementation of the German research program Modeling and Measuring Competencies in Higher Education (KoKoHs). The first phase of KoKoHs ran from 2011 to 2015. KoKoHs projects modeled domain-specific and generic competencies acquired at an academic level and transformed them into measurement models and instruments that were later tested empirically and validated (Zlatkin-Troitschanskaia, Pant, Kuhn, Toepper & Lautenbach, in press).

The KoKoHs research program pursued three general objectives:

1. to ensure and maintain the quality of the higher education system in Germany in the face of growing international competition;
2. to contribute to international research on competencies in higher education by ensuring international compatibility and visibility of KoKoHs research; and
3. to create a framework for evaluating the effectiveness of higher education to enable evidence-based policy decisions and institutional assessment.

Method

The KoKoHs program has attracted considerable research interest. Within KoKoHs, 24 alliances among researchers at almost 50 higher education institutions conducted 70 single projects. In each cross-university alliance, experts in the respective disciplines, educational studies, and measurement methodology also cooperated with national and international external partners. KoKoHs project teams took into account curricular and job-related requirements, transformed theoretical competency models into suitable measuring instruments, and validated test score interpretations. The first phase of KoKoHs focused on competencies in a number of large fields of study as well as on generic competencies such as self-regulation in higher education learning or in teaching practice as well as general research competencies including evidence-based argumentation in medicine and teaching or understanding scientific literature in educational and social sciences.

In line with an agreed upon holistic definition of competencies as latent cognitive and affective-motivational underpinnings of performance (see Weinert, 2001), KoKoHs researchers developed competency models and explored the methodology of competency assessment. This complex and multi-dimensional research area calls for complex and multi-dimensional research methods. Statements on the dimension, grading, and development of generic and domain-specific competencies are prerequisite to generating suitable measurement instruments. In KoKoHs, models of cognitive abilities and skills were operationalized through measuring instruments and tested in empirical assessments. Efforts were made to establish validity of the interpretation of the evidence so as to determine what can be inferred from the cognitive representations elicited by the assessment of the competencies of individual students.

The project teams conducted systematic, internationally compatible, fundamental research on theoretical modeling and empirical assessment of academic competencies of students in higher education, and they validated their test interpretations (Zlatkin-Troitschanskaia, Kuhn & Toepper, 2014). Although the KoKoHs program is based in Germany and focuses on student competencies in higher education at a national level, KoKoHs researchers have followed international best practices and standards in test adaptation and validation so as to enhance international comparability of assessments (Zlatkin-Troitschanskaia, Pant, Kuhn, Toepper & Lautenbach, in press).

We will discuss results from the first phase of KoKoHs in relation to an updated review of the state of international research (see Zlatkin-Troitschanskaia, Pant, Kuhn, Toepper & Lautenbach, 2016).

Results

In the KoKoHs research projects, more than 40 models of domain-specific and generic competencies were developed and validated. Furthermore, test instruments were developed or were adapted from existing international tests. Overall, approximately 60 paper-pencil tests and 40 computer-based instruments were developed and used to assess more than 50,000 students from approximately 220 institutions of higher education. These models of competency structures, assessment designs, and measuring instruments developed and tested in the projects provide a solid basis for higher education learning outcomes assessment in Germany. Results from the first five-year phase of KoKoHs indicate that the program generated evidence of the quality of developed models and instruments and of the reliability of information on the assessed competency constructs.

Reflection

The completion of the first phase of the KoKoHs research program marks an important milestone on the way towards valid and reliable assessment of academic competencies in Germany. However, the state of international research indicates that there are many challenges in valid assessment of learning outcomes still to be addressed.

In 2015, a new KoKoHs funding program on Validation and Methodological Innovations was initiated. Research in the new funding program must be based on precise descriptions of the competencies to be assessed and well-documented pilot studies evidencing the psychometric properties of the instruments. Preliminary studies may have been conducted in previous KoKoHs projects or externally. The competency models and corresponding assessment instruments will be used for in-depth field-experimental validation studies, including longitudinal and multilevel analyses.

In this new funding program, it will be essential to link and compare findings from Germany and other countries. As AHELO has shown, international comparative assessments are important and – despite many local and national differences – possible. For example, AHELO and current studies have adapted and used assessments of critical thinking skills in various countries. Further research on international comparisons based on such assessments could provide interesting insights and promote discussion not only on university admission, but also on preconditions of learning in higher education and their justification in a broad sense. With higher education being affected by internationalization and globalization, international cooperation in this field of research is gaining increasing importance.

References

Coates, H. (2014). *Higher Education Learning Outcomes Assessment – International Perspectives*. Frankfurt am Main: Peter Lang.

Kuhn, C., & Zlatkin-Troitschanskaia, O. (2011). *Assessment of Competencies among University Students and Graduates – Analyzing the State of Research and Perspectives* (Working Papers: business education, 59). Johannes Gutenberg University Mainz.

Land, R., & Gordon, G. (2013). *Enhancing Quality in Higher Education – International Perspectives*. London & New York: Routledge.

Liu, L. (2011). Outcomes Assessment in Higher Education: Challenges and Future Research in the Context of Voluntary System of Accountability. In *Educational Measurement: Issues and Practice, 30*(3), 2-9.

Tremblay, K., Lalancette, D., & Roseveare, D. (2012). *Assessment of Higher Education Learning Outcomes. Feasibility Study Report. Volume 1 – Design and Implementation*. OECD. http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf

Weinert, F. E. (2001). Concept of competence: A Conceptual Clarification. In Rychen, D. S. and Salganik, L. H. (Eds.), *Defining and Selecting Key Competencies* (pp. 45–65). Seattle, WA: Hogrefe and Huber.

Zlatkin-Troitschanskaia, O., Kuhn, C., & Toepper, M. (2014). Modeling and assessing higher education learning outcomes in Germany. In H. Coates (Ed.), *Higher Education Learning Outcomes Assessment – International Perspectives* (pp. 213-235). Frankfurt am Main: Lang.

Zlatkin-Troitschanskaia, O, Pant, H. A., Kuhn, C., Toepper, M., & Lautenbach, C. (2016). *Messung akademischer Kompetenzen von Studierenden und Hochschulabsolventen – Ein Überblick zum nationalen und internationalen Forschungsstand. [Measuring academic competencies of higher education students and graduates – An Overview of the state of national and international research.]* Wiesbaden: Springer.

Zlatkin-Troitschanskaia, O., Pant, H. A., Kuhn, C., Toepper, M., & Lautenbach, C. (in press). Assessment Practices in Higher Education and Results of the German Research Program Modeling and Measuring Competencies in Higher Education. *Research & Practice in Assessment*.

Acknowledgments
The KoKoHs program is funded by the German Federal Ministry of Education and Research.

Modelling factors that determine higher-education performance and estimate future educational outcomes

Eduardo C. Cascallar,
University of Leuven & Assessment Group International
Mariel F. Musso
University of Leuven & National Research Council (Argentina)
Eva Kyndt
University of Leuven

Abstract

Education has been impacted by the shift from an industrial society to an information-based environment. We are now shifting again to an "innovation-based" society which requires what Sternberg (2000) calls 'successful intelligence'. This is particularly true in the area of higher education, where outcome oriented reforms and pressure to obtain more valid information on students' outcomes, highlights the need to adequately model their performance and the factors that participate in those outcomes, while also understanding and being able to predict the results of the educational programmes. As the practice of educational assessment evolves, developments in cognitive science and psychometrics along with continuing advances in technology lead to new views of the nature and function of assessment (Segers, Dochy & Cascallar, 2003; Braun, 2005). New methodologies and technologies, and the emergence of predictive systems, have focused on the possibility of assessments which use a wide range of data or student productions to evaluate their performance without the need of traditional testing (Boekaerts & Cascallar, 2006). This research presents the application of educational assessments utilizing neural network predictive systems in three studies exploring models for general academic performance and performance in a specific field (mathematics). It introduces the application of these methodologies in education, and evaluates the results and quality of the predictive systems. Results from these methods achieved excellent levels of predictive classification, and facilitate the development of models that take into account cognitive, self-regulation and background factors in a comprehensive fashion, which takes into account all complex interactions. Their impact on the understanding of the processes involved, educational quality and improvement, as well as accountability is highlighted.

Key words: higher education, assessment, mathematics, neural networks, predictive systems.

Introduction

New applications have continuously been introduced which affect all aspects of the assessment process: knowledge base management, development of test items, computer delivery, and automated scoring. Currently, these advances cover a wide range of new applications using diverse technological advances, which result in the implementation of programs with novel technical and conceptual contributions. These new methodologies and technologies, and the emergence of predictive systems, have focused on the possibility of assessments which use a wide range of data or student productions to evaluate their performance without the need of traditional testing (Cascallar, Boekaerts & Costigan, 2006; Boekaerts & Cascallar, 2006). These new tools should be sensitive enough to accrue information about the level of performance that the students have reached so far in the domain of study. This approach should also include the prediction of the expected outcomes that best capture the students' current level of learning, using already available information.

Predictive streams analyses (Cascallar, Boekaerts & Costigan, 2006; Cascallar & Musso, 2008), based in this case on neural network (NN) models, have several strengths: (a) because these are machine learning algorithms, the assumptions required for traditional statistical predictive models (e.g., ordinary least squares regression) are not necessary. As such, this technique is able to model

nonlinear and complex relationships among variables. NNs aim to maximize classification accuracy and work through the data in an interactive process until maximum accuracy is achieved, automatically modeling all interactions among variables; (b) NNs are robust, general function estimators. They usually perform prediction tasks at least as well as other techniques and sometimes perform significantly better (Marquez, Hill, Worthley & Remus 1991); (c) NNs can handle data of all levels of measurement, continuous or categorical, as inputs and outputs. Because of the speed of microprocessors in even basic computers, NNs are more accessible today than they were when originally developed.

The NN learns by examining individual training case, then generating a prediction for each testing case, and making adjustments to the weights whenever it makes an incorrect prediction. Information is passed back through the network in iterations, gradually changing the weights. As training progresses, the network becomes increasingly accurate in replicating the known outcomes. This process is repeated many times, and the network continues to improve its predictions until one or more of the stopping criteria have been met. A minimum level of accuracy can be set as the stopping criterion, although additional stopping criteria may be used as well (e.g., number of iteration, amount of time). Once trained, the network can be applied to future cases (validation or holdout sample) for validation and implementation (Lippman, 1987).

Neural networks in educational research

NNs have been used in several different fields of research and in applied environments, such as: biology, business, finance, medicine, meteorology, environmental studies, and in the prediction of terrorist attacks, among other applications. During the last few decades, NNs have been increasingly utilized as a statistical methodology in applied areas such as classification and recognition of patterns in business and the social sciences (Al-Deek, 2001; Neal & Wurst, 2001; Nguyen & Cripps, 2001; White & Racine, 2001; Laguna & Marti, 2002; Detienne, Detienne & Joshi, 2003). However, the literature shows very few studies applying neural networks in education and in educational assessment in particular (Wilson & Hardgrave, 1995), even though some authors have called attention to the fact that traditional statistical methods do not always yield accurate predictions and/or classifications (Everson, 1995). Preliminary research applying artificial intelligence computing methods to problems of prediction, selection and classification (Perkins, Gupta & Tammana, 1995) suggests that artificial neural networks and other neural computing methods may substantially improve the validity of the classifications, as well as increase the accuracy of classifications, and also improve the predictive validity of test scores and other educational information (Everson, Chance, & Lykins, 1994). Another study (Hardgrave, Wilson, & Walstrom, 1994) compared a neural network model to other techniques in predicting graduate student success. They evaluated five different models: least squares regression, stepwise regression, discriminant analysis, logistic regression, and neural networks. Results of their study showed that neural networks "perform at least as well as traditional methods and are worthy of further investigation" (p. 249). Similarly, Gorr (1994) used neural networks to model the decision-making process of college admissions. Neural networks were compared with linear regression, stepwise polynomial regression, and an index used by the graduate admissions committee. These researchers found that "…a neural network identifies additional model structures over the regression models" (p. 17), and that even though a neural network model can address some of the same research issues as a conventional regression, a neural network is inherently a different mathematical approach (Detienne, Detienne & Joshi, 2003). In terms of their application, neural networks have been considered to be especially good as statistical models when the emphasis is on prediction and/or classification of complex phenomena and recent developments have provided tools to look into the "black box" of the network, and to shed light on the interrelationships of the variables involved in the network calculations.

Three Research Studies in Higher Education

Using this approach, three studies conducted in the area of higher education modeled general academic outcomes (Study 1 and 2), and performance in mathematics (Study 3).

Study 1: A total sample of 864 university students of both genders, ages ranging between 18 and 25 was used. Three neural network models were developed. Two of the models (identifying the top 33% and the lowest 33% groups, respectively) were able to reach 100% correct identification of all students in each of the two groups. The third model (identifying low, mid and high performance levels) reached precisions from 87% to 100% for the three groups. Analyses also explored the predicted outcomes at an individual level, and their correlations with the observed results, as a continuous variable for the whole group of students. Results demonstrate the greater accuracy of the ANN compared to traditional methods such as logistic regression. In addition, the ANN provided information on those predictors that best explained the different levels of expected performance (cognitive factors in the "low" group, and self-regulation and background factors in the "high" group (Musso, Kyndt, Cascallar & Dochy, 2013).

Study 2: In this study three neural network analyses were performed for three categories of academic performers: the top 20%, the bottom 20%, and the 60% middle group of students. Participants in this study were 128 university students. Precisions were 90%-95% in all models. Results show that working memory capacity and attention are both good predictors of academic performance especially for the best and the weakest performers of the group. Students' motivation and approaches to learning were good predictors for the group of students whose performance was in the middle 60% (Kyndt, Musso, Cascallar & Dochy, 2015).

Study 3: A total sample of 800 entering university students of both genders, ages ranging between 18 and 25 was used. Three neural network models were developed to identify the lowest 30%, highest 30%, and middle 30% group of students, respectively, in terms of their estimated future performance in a mathematics test. Two of the models (identifying the top 30% and the low 30% groups) were able to reach 100% correct identification of all students in each of the two groups, using the corresponding ANN. The third model (middle 30% group) was able to reach 70% correct identification. These ANN models showed interesting differences in the pattern of relative predictive weight importance amongst those variables with the highest participation for the predictive model. For "low" performers, basic cognitive variables were most important, while self-regulation and background variables were good predictors for "high" performers (Musso, Kyndt, Cascallar & Dochy, 2012).

In all three studies, the ANN models used were backpropagation multilayer perceptron neural networks, that is, a multilayer network composed of nonlinear units, which computes its activation level by summing all the weighted activations it receives and which then transforms its activation into a response via a nonlinear transfer function. After an initial training phase in which the model was developed, in each case a randomly selected set of data from the same dataset was used to test and validate the results of the training phase. Results indicated a complex pattern of interactions between the various sources of data, while indicating the individual contribution of each variable (as well as set of variables) for each level of academic performance. These models shed light on the participation of factors in the attainment of levels of academic outcomes in higher education. In addition, they suggest the areas in which potential interventions could improve such performance. Applications of this approach range from the study of sources of variance in academic outcomes, selection, retention, diagnostic assessment, and prediction of expected outcomes under certain environmental and student conditions. Implications for the conceptualization of new modes of assessment without the use of testing or questionnaires, as well as the role of various factors in academic outcomes in higher education are explored.

References

Al-Deek, H. M. (2001). Which method is better for developing freight planning models at seaports – Neural networks or multiple regression? *Transportation research record, 1763*, 90- 97.

Boekaerts, M., & Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in Self-regulation? *Educational Psychology Review*, *18*(3), 199-210.

Braun, H. (2005). Value-added modeling: What does due diligence require. Value added models in education: Theory and practice, 19-38.

Cascallar, E. C., Boekaerts, M., & Costigan, T. E. (2006) Assessment in the Evaluation of Self- Regulation as a Process, *Educational Psychology Review*, *18*(3), 297-306.

Cascallar, E. C., & Musso, M. F. (2008). Classificatory stream analysis in the prediction of expected reading readiness: Understanding student performance. *International Journal of Psychology, Proceedings of the XXIX International Congress of Psychology ICP 2008, 43*(43/44), 231-.231.

Detienne, K. B., Detienne, D. H., & Joshi, S. A. (2003). Neural networks as statistical tools for business researchers. *Organizational Research Methods, 6,* 236-265.

Everson, H. T. (1995). Modelling the student in intelligent tutoring systems: The promise of a new psychometrics. *Instructional Science, 23*(5-6), 433-452.

Everson, H. T., Chance, D., & Lykins, S. (1994). *Exploring the use of artificial neural networks in educational research.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Gorr, W. L. (1994). Research prospective on neural network forecasting. *International Journal of Forecasting*. *10* (1), 1-4.

Hardgrave, B. C., Wilson, R. L., & Walstrom, K. A. (1994).Predicting Graduate Student Success: A Comparison of Neural Networks and Traditional Techniques. *Computer and Operations Research, 21*(3), 249-263.

Kyndt, E., Musso, M, Cascallar, E. and Dochy, F. (2015). Predicting academic performance: The role of cognition, motivation and learning approaches. A neural network analysis. In V. Donche, S. De Maeyer, D. Gijbels, & H. van den Bergh (Eds). Methodological Challanges in Research on Student Learning. Garant: Antwerp, Belgium

Laguna, M. & Marti, R. (2002).Neural network prediction in a system for optimizing simulations. *IIE Transactions, 34* (3), 273-282.

Lippman, R. (1987). An introduction to computing with neuralets. *IEEE ASSP Magazine, 3*(4), 4-22.

Marquez, L., Hill, T., Worthley, R., & Remus, W. (1991). Neural network models as an alternative to regression. *Proceedings of the IEEE 24th Annual Hawaii International Conference on Systems Sciences, 4,* 129-135.

Musso, M. F., Kyndt, E., Cascallar, E. C., & Dochy, F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research*, *1*(1). doi:10.14786/flr.v1i1.13

Musso, M., Kyndt, E., Cascallar, E., & Dochy, F. (2012). Predicting Mathematical Performance: The Effect of Cognitive Processes and Self-Regulation Factors. *Education Research International,* 2012, 1-13. doi:10.1155/2012/250719

Neal, W. & Wurst, J. (2001).Advances in Market Segmentation. Marketing Research, 13(1), 14-18.

Neal, W., & Wurst, J. (2001).Advances in market segmentation. *Marketing Research, 13*(1), 14-18.

Nguyen, N. & Cripps, A. (2001).Predicting housing value: A comparison of multiple regression and artificial neural networks. *Journal of Real Estate Research, 22*(3), 313-336.

Perkins, K., Gupta, L. & Tammana (1995). Predict item difficulty in a reading comprehension test with an artificial neural network. *Language Testing, 12*(1), 34-53.

Sternberg, R. J. (2000). Identifying and developing creative giftedness. *Roeper Review*, *23*(2), 60-64.

Segers, M., Dochy, F., & Cascallar, E. (2003).Optimizing new modes of assessment: In search of qualities and standards. The Netherlands: Kluwer Academic Publishers.

White, H. & Racine, J. (2001): Statistical inference, the bootstrap, and neural network modelling with application to foreign exchange rates. *IEEE Transactions on Neural Networks: Special Issue on Neural Networks in Financial Engineering, 12,* 657-673.

Wilson, R. L. & Hardgrave, B. C. (1995). Predicting graduate student success in a MBA program: Regression vs. classification. *Educational and Psychological Measurement, 55,* 186-195.

Evaluating stability of self-reported personal dispositions: A repeated measures study of undergraduate students

Gavin T L Brown & Makayla P. Grays
University of Auckland

Abstract: Higher education learning outcomes include a range of personal dispositions or attributes that are conventionally evaluated by collected self-reported responses to questionnaire items intended to measure latent traits. Important personal traits sought by universities include intellectual curiosity and openness to diverse ideas, experiences, and peoples. Evaluators are reliant on robust social psychological instruments as measures. However, the validity and stability of such instruments across repeated administrations needs empirical evidence. This paper reports a repeated measures (early and late the same academic year) factor analytic study using three cohorts of students (first-year undergraduate, final-year undergraduate, and graduates) in one faculty to evaluate the psychometric properties of scales focused on curiosity and openness to diversity. Confirmatory factor analysis with invariance testing showed that the measurement models developed at Time 1 were not well-fitting at Time 2. Re-analysis of data from both times identified revised models which had configural, metric, and scalar invariance across both time points. This study points to difficulties in obtaining stable estimates of self-reported psychological traits and need to evaluate data carefully.

Introduction

There is reasonable agreement that university graduates should have a range of competences beyond discipline specific skills. In New Zealand and Australia these competences are known as 'graduate attributes', which are "the skills, knowledge and abilities of university graduates, beyond disciplinary content knowledge, which are applicable in a range of contexts and are acquired as a result of completing any undergraduate degree" (Barrie, 2006, p. 217). These competences can be aggregated into three major groups (i.e., conceptual, personal, and people skills) (Strijbos, Engels, & Struyven, 2015). A number of non-cognitive personal dispositions or skills have been found to be associated with positive attitudes and academic success. These are generally measured indirectly relying on self-reported self-perceptions and self-evaluations (Zlatkin-Troitschanskaia et al., 2015).

The two dispositions of interest in this study are intellectual curiosity and openness to diversity. Curiosity, which is a desire "to acquire new knowledge, including beliefs or feelings of surprise, intrigue, and incomplete information about a topic" has been found to be a positive motivator for persistence in academic settings (French & Oakes, 2003, p.89). Curiosity also captures the idea of people wanting to 'stretch' their capabilities by actively seeking out new information or experiences curiosity and has shown positive relations with effective reappraisal skills, willingness to express positive emotions, and the ability to persist at goal-directed behavior (Kashdan et al., 2007).

With increasing diversity in enrolment, student interactions with people from different backgrounds leads to greater understanding and appreciation of human diversity (Kuh, et al., 2003). Greater enjoyment in being intellectually challenged by different ideas, values, and perspectives, as well as an appreciation of racial, cultural, and value diversity, seem to be hallmarks of a positive experience at university (Kuh et al., 2003). Ethnocultural empathy involves the ability to understand and pay attention to the feelings of "people of racial and ethnic backgrounds different from one's own" (Wang et al., 2003, p. 221) and is associated with prosocial behaviour. The ability to distinguish important human differences based on differential membership of cultural, ethnic, linguistic, etc. groups combined with the ability to perceive shared commonalities despite differences in membership is associated with positive attitudes towards diversity of people in higher education, as well as constructive coping skills (Fuertes, Miville, Mohr, Sedlacek, & Gretchen, 2000). Hence, successful students seem to be open to diversity.

Research Question

In order to make valid comparisons between groups and across times the survey instrument must elicit statistically equivalent responding or measurement invariance (Cheung & Rensvold, 2002) so that comparison of mean scores between times or groups is not confounded by differential responding to the survey items. Measurement invariance requires that a model has configural equivalence (accepted if RMSEA value is <.05), that regression weights from the factor to each item (i.e., metric equivalence) are equivalent (accepted if ΔCFI is ≤.01), and that intercepts of each item on respective factors (i.e., scalar equivalence) are equivalent (also accepted if ΔCFI is ≤.01). Hence, this study focuses on determination of measurement models that are statistically equivalence for both times of administration.

Method

Following conventional scientific practice, scales and items in the published research literature were selected to stimulate student responses. A quasi-experimental, repeated-measures design tested for measurement invariance.

Participants

The Early (*n*=339) and Late 2014 (*n*=165) surveys had mostly female (about 80%) and European ethnicity (45% Early, 55% Late) participants. The proportions of first-year (35%), final-year (32%), and Graduate Diploma (GradDip) (29%) participants were comparable in Early 2014. However, in Late 2014, participation was higher among first-year (48%) than final-year (23%) or GradDip (25%) students. Of these, 39 1st-year students did both surveys, as did 34 final years, and 40 GradDips.

Instruments

The questionnaire was analysed in two conceptual parts; Part A included 20 items related to intellectual openness and curiosity, as well as love and enjoyment of ideas, discovery and learning, while Part B included 30 items related to openness to diversity.

Part A. Four items were selected with modifications from the Academic Intrinsic Motivation Scale (AIMS) (French & Oakes, 2003). One item from the Curiosity and Exploration Inventory-II (CEI-II) (Kashdan et al., 2009) *stretching* scale (i.e., motivation to seek out knowledge and new experiences) was used. Nine items from the Melbourne Curiosity Index (MCI) (Naylor, 1981) were selected because they represented attributes of curiosity and love of learning.

Part B. Five items were adapted from the College Student Experience Questionnaire (CSEQ) (Kuh et al., 2003) because of their strong academic focus (i.e., referring to diverse and challenging experiences in students' courses) and alignment with the attribute openness to diversity. Five items representing a focus on ethnic/cultural/demographic diversity from the Miville-Guzman Universality-Diversity Scale (MGUDS) (Fuertes et al., 2000) were selected and modified. Just one item from the Scale of Ethnocultural Empathy (SEE) (Wang et al., 2003) was considered appropriate for inclusion.

Response Format. University students and graduates because of their commitment to learning were expected to be positively inclined towards the dispositions. Consequently, a 6-point positively packed response scale was used with the following response options and score values: (1) *Strongly disagree*, (2) *Mostly disagree*, (3) *Slightly agree*, (4) *Moderately agree*, (5) *Mostly agree*, (6) *Strongly agree*.

Results

Intellectual curiosity

The 15-item, correlated two-factor model from Early 2014 did not fit the Late 2014 data. A new model was developed using both data sets and tested for fit in both Early (*n*=310) and Late 2014 (*n*=153) samples. Various one-, two- and three-factor solutions were tested and overall fit improved by alternating items in each solution. A 15-item, correlated three-factor model had adequate fit to both the Early (χ^2/df=2.59, RMSEA=.07, CFI=.96, SRMR=.03) and Late 2014 (χ^2/df=2.04, RMSEA=.08, CFI=.94, SRMR=.05) data. Metric and scalar invariance were accepted with change in CFI <.01. The

item content was used to label the three factors/scales as: (1) Curiosity, (2) Love of learning, and (3) Answer-seeking (Figure 1, Panel A).

Openness to Diversity

A 7-item, one-factor model from Early 2014 did not fit the Late 2014 sample. A new model using both Early (*n*=302) and Late 2014 (*n*=147) samples was developed. Various one-, two-, three- and four-factor solutions were tested and overall fit improved by alternating items in each solution. A 12-item, three-factor model yielded good fit in both Late 2014 (χ^2/df=2.04, RMSEA=.08, CFI=.95, SRMR=.04) and Early 2014 (χ^2/df=2.82, RMSEA=.08, CFI=.96, SRMR=.04). Metric and scalar invariance were accepted with change in CFI <.01. The item content led to naming the three factors/scales Openness to diverse (1) perspectives/ideas, (2) cultures/groups, and (3) background/ individuals (Figure 1, Panel B).

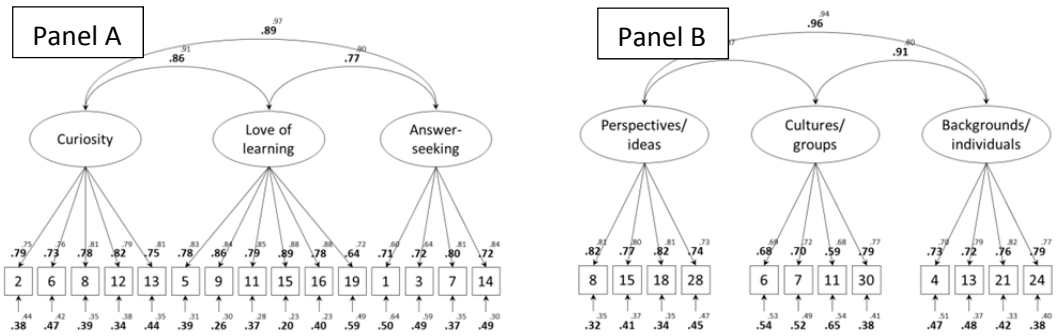

Figure 1. Standardised pattern coefficients and error variances for Intellectual Curiosity (Panel A) and Openness to Diversity (Panel B) in Late 2014 with Early 2014 estimates superscripted.

Discussion

The current measurement models for intellectual curiosity and openness to diversity demonstrated statistical equivalence across the two times of measurement, but only after both times of administration were available. This suggests HELO evaluation studies are very contingent on samples available for study and that careful analysis is needed before substantive claims can be made. Obtaining high response rates and large sample sizes is essential and without which claims are dubious.

References

Barrie, S. C. (2006). Understanding what we mean by the generic attributes of graduates. *Higher Education, 51*, 215–241. doi:10.1007/s10734-004-6384-7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

French, B. F., & Oakes, W. (2003). Measuring academic intrinsic motivation in the first year of college: Reliability and validity evidence for a new instrument. *Journal of The First-Year Experience, 15*(2), 83-102.

Fuertes, J. N., Miville, M. L., Mohr, J. J., Sedlacek, W. E., & Gretchen, D. (2000). Factor structure and short form of the Miville-Guzman Universality-Diversity Scale. *Measurement and Evaluation in Counseling and Development, 33*, 157-169.

Kashdan, T. B., Gallagher, M. W., Silvia, P. J., Winterstein, B. P., Breen, W. E., Terhar, D., & Steger, M. F. (2009). The Curiosity and Exploration Inventory-II: Development, factor structure, and psychometrics. *Journal of Research in Personality, 43*, 987-998. doi:10.1016/j.jrp.2009.04.011

Kuh, G. D., Gonyea, R. M., Kish, K. E., Muthiah, R., & Thomas, A. (2003). College Student Experiences Questionnaire: Norms for the 4th Edition. . Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.

Strijbos, J., Engels, N., & Struyven, K. (2015). Criteria and standards of generic competences at bachelor degree level: A review study. *Educational Research Review, 14*, 18-32. doi:10.1016/j.edurev.2015.01.001

Wang, Y.-W., Davidson, M. M., Yakushko, O. F., Savoy, H. B., Tan, J. A., & Bleier, J. K. (2003). The Scale of Ethnocultural Empathy: Development, validation, and reliability. *Journal of Counseling Psychology, 50*(2), 221-234. doi:10.1037/0022-0167.50.2.221

Zlatkin-Troitschanskaia, O., Shavelson, R. J., & Kuhn, C. (2015). The international state of research on measurement of competency in higher education. *Studies in Higher Education, 40*(3), 393-411. doi:10.1080/03075079.2015.1004241

Acknowledgments

This research was funded by the Vice-Chancellor Strategic Development Fund Project #23602.