

Multilevel Modelling of Learning Gains: The Impact of Module Particulars on Students' Learning in Higher Education. (0215)

Jekaterina Rogaten¹, Bart Rienties¹, Denise Whitelock¹, Simon Cross¹, Allison Littlejohn¹, Rhona Sharpe², Simon Lygo-Baker³, Ian Scott², Steven Warburton³, Ian Kinchin³

¹The Open University UK, UK

²Oxford Brooks University, UK

³University of Surrey, UK

Introduction

Education researchers and practitioners have been developing and testing a range of measurement approaches aiming to capture relative improvements in each individual student's learning, rather than their performance relative to the type of 'comparative measure' (e.g., Cahill et al., 2014; Hake, 1998; Mortensen & Nicholson, 2015). This relative measure gives critical insight into the progress each student makes while learning. One approach commonly used is termed 'learning gains', which can be defined as growth or change in knowledge, skills, and abilities over time that are linked to the desired learning outcomes or learning goals of the course (e.g., Cronbach & Furby, 1970; Linn & Slinde, 1977; Lord, 1956, 1958).

The most common way of assessing learning gains is through use of pre- post testing (e.g., Harris, 1963; Lord, 1956, 1958). However, rolling out this type of testing across an entire university with a wide variety of disciplines is complex, and in some cases, not possible due to a substantial variation in learning objectives between disciplines. One possible solution is to use students' academic performance as a proxy for estimating learning gains. This approach capitalises on the large quantity of student data routinely gathered by every university and, at the same time, offers opportunity to measure learning gains across various disciplines and even across different universities.

Learning gains can be calculated in a number of ways. The most common approach is to compute the normalised learning gain on pre-post test scores (Hake, 1998). However, normalised learning gains represent an accurate estimate of learning only when students perform better in post-tests than in pre-tests (Marx & Cummings, 2007), which is not always the case when looking at academic performance in Higher Education (Jensen, Kummer, & Godoy, 2015; Yalaki, 2010). Furthermore, normalised learning gains are only possible to compute on two set of scores, and if there are more than two observations, other statistics like

ANOVA and ANCOVA should be used as they allow comparison between more than two data points (Dimitrov & Rumrill Jr, 2003). However, a limitation of this method is that general linear models are based on the assumption that observed data of one student are independent of observed data of another. This assumption is not true in situations where students from the same discipline, identical module, and same university have similar experiences (e.g., easy assessment in week 1, difficult assessment in week 6, and moderately hard final assessment in week 15) which means the variance in their scores are not independent and are linked. This means that simple linear models to measure learning gains may provide inaccurate and misleading outcomes.

The aim of this research is to establish whether academic performance within modules is a valid proxy for estimating students' learning gains, and whether there is variance in learning gains that is due to students having shared educational experiences at the level of a module.

Method

A total of 21,192 undergraduate degree students were sampled from an Open University UK dataset. There were 10,038 STEM students of whom 74.4% were males, 25.6% were females, with average age of 30 years. In addition, 11,154 Business and Law students were included, of whom 45.3% were males, 54.7% were females, with average age of 32 years. Academic performance for each tutor marked assessment within each of the 200 modules was retrieved from our university database for 2012/13, 2013/14 and 2014/15 academic years.

The data were analysed using three-level linear growth-curve model estimated in MLWiN (Rasbash, Charlton, Browne, Healy, & Cameron, 2005; Rasbash, Steele, Browne, & Goldstein, 2009). Identical models were estimated for each of the two faculties. Level 1 was 'tutor marked assessments' that students completed throughout the module. Level 2 was 'student' and level 3 was 'module'. The dependent variable was student academic performance on each of the tutor marked assessments with the possible maximum score of 100.

Results

Business and Law students' academic achievements were on average $M = 64$; $SD = 15.2$. The results of the growth curve model estimation illustrated that 'module' accounted for 16% of variance in students' initial achievements, and 51% of variance in subsequent learning gains. Students with initial high achievements showed lower learning gains ($r = -.1$); however the correlation was weak. A module level intercept-slope correlation was much stronger ($r = -.43$), indicating that most variance was accounted for by differences between modules.

STEM students' academic achievement was on average $M = 71.6$; $SD = 21.4$. The multilevel modelling showed that 'module' accounted for 46% of variance in initial achievements, and 77% of variance in subsequent learning gains. Similar to the Business and Law example, STEM students with initial high achievements also showed lower learning gains ($r = -.15$), whereas module level intercept-slope correlation was ($r = -.69$), thus most of the variance was accounted for by differences between modules. Therefore, in both faculties students who studied in modules with initial high student achievements showed lower learning gains than students who studied in the modules with low initial student achievements.

Discussion

Overall this research has two key findings which have important theoretical and practical implications for the measurement of learning gains in Higher Education. Firstly, the results illustrate that the specific module that a student is enrolled in accounts for a substantial portion of variance, not just in the student's initial academic achievements, but also in the learning gains throughout the module. While there are some modules where students showed positive learning gains, other modules showed negative learning gains and students who started with high initial achievements demonstrated low learning progress. As such, multilevel modelling is a more accurate method compared with simple linear models to estimate students' learning gains. The simple models are not able to detect differences between modules when looking at the faculty level performance whereas multilevel modelling does. This has important implications for TEF as when assessing learning gains on an institutional level, looking at the whole institution or faculty or department performance can result in misleading estimate of students' learning gains.

Secondly, low learning gains - or negative learning gains of high achieving students - does not imply that the students are losing knowledge or ability *per se*. However it highlights

the complexity of factors that have to be taken into account when using students' academic performance as a proxy for learning gains. These factors include 'assessment difficulty' and 'learning design' (Rienties & Toetnel, 2016). Examination of these specific factors was outside the scope of this study, but future research will examine whether adjustments can be made to using academic performance as a proxy for more accurate estimates of learning gains within Higher Education.

References

- Cahill, M. J., Hynes, K. M., Trousil, R., Brooks, L. A., McDaniel, M. A., Repice, M., ... Frey, R. F. (2014). Multiyear, multi-instructor evaluation of a large-class interactive-engagement curriculum. *Physical Review Special Topics-Physics Education Research*, *10*(2), 020101–020119.
- Cronbach, L. J., & Furby, L. (1970). How we should measure 'change': Or should we? *Psychological Bulletin*, *74*(1), 68–80.
- Dimitrov, D. M., & Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work*, *20*(2), 159–165.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74.
- Harris, C. (1963). *Problems in Measuring Change*. UNIV OF WI+PRESS.
- Jensen, J. L., Kummer, T. A., & Godoy, P. D. D. M. (2015). Improvements from a Flipped Classroom May Simply Be the Fruits of Active Learning. *Cbe-Life Sciences Education*, *14*(1), 71-76.
- Linn, R. L., & Slinde, J. A. (1977). The Determination of the Significance of Change between Pre- and Posttesting Periods. *Review of Educational Research*, *47*(1), 121–150.
- Lord, F. M. (1956). The Measurement of Growth. *ETS Research Bulletin Series*, *1956*(1), i-22.

- Lord, F. M. (1958). Further problems in the measurement of growth. *Educational and Psychological Measurement, 18*, 437–451.
- Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75*(1), 87–91.
- Mortensen, C. J., & Nicholson, A. M. (2015). The flipped classroom stimulates greater learning and is a modern 21st century approach to teaching today's undergraduates. *Journal of Animal Science, 93*(7), 3722–3731.
- Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2005). MLwiN Version 2.02. (Version 2.02). Centre for Multilevel Modelling: University of Bristol.
- Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2009). A User's Guide to MLwiN. *Centre for Multilevel Modelling, University of Bristol, 296*.
- Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior, 60*, 333–341.
- Yalaki, Y. (2010). Simple Formative Assessment, High Learning Gains in College General Chemistry. *Egitim Arastirmalari-Eurasian Journal of Educational Research, 10*(40), 223–243.