Assessing Learning Gains

Jekaterina Rogaten, Bart Rienties and Denise Whitelock

Open University, Milton Keynes, UK
Jekaterina.rogaten@open.ac.uk

Abstract. Over the last 30 years a range of assessment strategies have been developed aiming to effectively capture students' learning in Higher Education and one such strategy is measuring students' learning gains. The main goal of this study was to examine whether academic performance within modules is a valid proxy for estimating students' learning gains. A total of 17,700 Science and Social Science students in 111 modules at the Open University UK were included in our three-level linear growth-curve model. Results indicated that for students studying in Science disciplines modules, module accounted for 33% of variance in students' initial achievements, and 26% of variance in subsequent learning gains, whereas for students studying in Social Science disciplines modules, module accounted for 6% of variance in initial achievements, and 19% or variance in subsequent learning gains. The importance of the nature of the consistent, high quality assessments in predicting learning gains is discussed.

Keywords: learning gains · grades · assessment · multilevel modelling · Higher Education.

1 Introduction

Over the years a variety of assessment strategies have been developed aiming to effectively capture students' learning in Higher Education (HE) [1]. Throughout HE sector universities are using summative assessment, but there is now an increasing number of institutions which are using Computer Based Assessment (CBA) to deliver, monitor, and evaluate students' learning [2–4]. The feedback students receive from CBA is often limited to a grade [2, 5], however there is also formative CBA that is used to inform students and educators of learning progress [6]. Information provided by formative CBA can help to shape learning, and is particularly useful when it is available to learners either before they start work or during the learning process [2, 5, 7, 8].

Given the near universal nature of assessing students' learning in HE, several researchers have used assessment results as proxies for learning gains, which are defined in this article as the change in knowledge, skills, and abilities over time as a result of targeted learning process [9–12]. There are multiple learning gains that students can develop in HE, which are linked to the learning outcomes or learning goals of the course: development of the conceptual understanding of the topic [13]; scientific reasoning and confidence in reasoning skills [14]; scientific writing and reading [15]; critical thinking [16]; problem solving, creativity, analytical ability, technical skills and communication [17]; moral reasoning [18]; leadership [19]; interest in political and social environment [20]; well-being [21]; and motivation [22]. Measuring such a variety of learning gains is a challenge in itself and a number of methodologies have been used to assess them. The approaches range from pre-post testing using standardised tests to cross-sectional studies using self-reported measures. Assessment of learning gains in knowledge and understanding is no exception and different methods are routinely used.

For example, Hake [13] examined students' learning gains in conceptual understanding of Newtonian mechanics in a sample of 6,542 undergraduate students using standardized tests at the beginning (pre-test) and at the end (post-test) of the course. Similar studies were undertaken by other teams of researchers [23, 24] who also used standardised test in the beginning and end of a semester to capture students' learning gains. These studies reported students making low to moderate learning gains during the limited time of one semester.

A recent meta-analysis by Rogaten and colleagues [25] amongst 51 learning gains studies indicated that with regards to cross-sectional studies, knowledge and understanding along with other learning gains were most often measured with Student Assessment of Learning Gains (SALG) scale [17, 26, 27] . SALG is a self-reported questionnaire that assesses students' perceived level of learning gains. There are also other measures that can be used to reliably assess students' perceptions of learning gains in knowledge and understanding, such as Science Students Skills Inventory (SSSI) [28, 29], and Student Experience in the Research University Survey (SERU-S) [30]. Since these instruments use self-reported measures, these type of studies rely on the premise that students can realistically and adequately appraise their own learning gains, which of course can be disputed [31].

The use of objective tests and pre-post testing to capture students' learning is generally preferred over the use of self-reported measures. Objective tests may capture unbiased learning gains rather than the perceptions of learning gains, and therefore are less reliant on individuals' abilities to properly self-appraise their own learning progress. However, pre-post testing is more resource-intensive in comparison to administration of self-reported surveys at the end of modules, and may become even more cost-intensive if teachers, universities, and governments want to estimate learning gains across various disciplines and number of universities [32].

A potential alternative to the administration of pre-post tests for assessing students' gains in knowledge and understanding is to estimate students' learning gains from course assessments grades. This approach capitalises on the large quantity of student data routinely gathered by every university and, at the same time, offers opportunities to measure learning gains across various disciplines and universities without additional measurement and financial costs. Furthermore, using students' academic performance as a measure of learning progress has other advantages; firstly, it is widely recognized as an appropriate measure of learning, secondly, it is relatively free from self-reported biases, and thirdly, using academic performance allows a direct comparison of research finding with the results from other studies [33–35].

At the same time, using academic performance scores as proxies for learning might have several limitations, such as a lack of assessment quality (e.g., too easy or too hard, focused on knowledge reproduction rather than critical evaluation) [2, 36], low inter-rater reliability (i.e., two markers give different assessment scores), and/or lack of coherence of assessment difficulty throughout the module (e.g., hard first assessment and easy final assessment; simple first assessment, hard second assessment, easy final assessment) [37, 38]. Therefore, in this article we will ask the following two research questions:

1. To what extent do assessment scores provide a valid, reliable proxy of estimating students' learning gains
2. How much variance in students' learning gains is accounted for by assessments, module characteristics and socio-demographic factors (i.e., gender, ethnicity and prior educational experience)?

In this study, we will use a three-level growth-curve model estimated for 17,700 HE students studying in two distinct disciplines (Science and Social science) in 111 modules at the Open University UK. After a brief review of assessment and feedback literature, we will review how researchers have used assessments as proxies for learning gains.

1.1 Importance of assessment and feedback.

The majority of HE institutions use assessment and feedback as a driver for and of learning. CBA has a lot of potential applications [4, 39] and benefits are being realized. There are a number of definitions and applications of CBA, but in the context of this study we conceptualize CBA as assessment presented using digital means and submitted electronically. CBA has numerous advantages [40] when compared to other, more traditional types of assessments. The most relevant benefits in distance-learning settings include more authentic interactive assessment options, such as intelligent tutoring [41], authentic virtual labs [42], speed of assessment, automatic feedback [43], and record-keeping. Although CBA is often used for summative assessments to evaluate what students learned, there has been an increase in use of CBA as a formative assessment in a form of online practice quizzes, wikis and peer assessment to provide formative feedback for students [2, 7, 44–46]. Using CBA for summative assessment only provides feedback in a form of a grade once all learning activities are completed [2, 5], whereas using CBA for formative assessment provides information that can help to shape learning, and is particularly useful when it is available to learners either before they start work or during the learning process [6]. As such, CBA is a valuable tool for helping students to regulate their learning processes [2, 5, 7, 8] .

A vast body of research has indicated that providing feedback is more important for learning than the assessment of learning [7]. Feedback originates from the field of engineering and information theory with the general assumption that information about the current system's state is used to change the future state. In his meta-study of 800+ meta-studies, Hattie [7] found that the way in which students receive feedback is one of the most powerful factors associated with the enhancement of learning experiences. Hattie and Yates [47](p. 60) consider feedback as empowering because it enables the learner to "move forward, plot, plan, adjust rethink and exercise self-regulation". For example, Whitelock [48] has argued that feedback is rather restrictive in nature when formative assessment's focus is that of "Assessment for Learning". She suggests that what is required in this context is a concept known as "Advice for Action". This approach does not restrict itself to giving advice after a task has been completed but can also embrace hints given before an assessment task is taken up.

1.2 Measuring and computing learning gains.

In the field of learning gains research, only a couple of studies have estimated learning gains from students' academic performance [49, 50] and overall students showed on average a decrease in their grades from the first to the last assessment of a semester/course. For example, Jensen and colleagues [51] assessed 108 biology course students and used results of 3 interim exams to estimate students' learning gains. Although in their study they focused on how students differed in flipped and non-flipped classrooms in terms of their academic performance across assessments, they reported that over the three unit exams students' performance generally decreased from 81.7% to 75.9% in non-flipped classroom and from 82.4% to 76.3% in flipped classroom. Thus, this decrease was equivalent in both groups of students. Yalaki [50] similarly assessed 168 organic chemistry students using their performance on 2 mid-term examinations and final exam. The goal of this study was to compare whether formative assessments and feedback resulted in better students' attainments in comparison to no formative assessment. They found that performance gradually decreased from the first interim exam to the final examination result (i.e., from 87.3% to 68.8% for group receiving formative assessments and feedback, and from 66% t0 61.4% for group receiving no formative assessments). In both of these studies researchers did not examine students' learning gains per se, but rather were interested in group differences in attainment on any one assessment. However, the observed decrease in attainments throughout the semester is contrary to what was found in pre-post test studies using standardized tests [13, 23, 24], where student on average showed an increase in their knowledge and understanding.

In addition to using different means to assess students' learning gains that seem to provide different results, there are a number of ways to compute students' learning gains [9, 11, 13, 52–54]. On the one hand, if one wants to examine the level of knowledge students developed over a course, one would assume that subtracting the beginning of a semester knowledge test score from the end of a semester knowledge test score will produce an accurate level of change/gain in academic achievement. Although this computation of learning gain makes intuitive sense, raw gain as a value of gain is inaccurate due to the difference between scores being less reliable than scores themselves [11], thus, it does not account for random error of measurements between pre-test and post-test scores [9, 10, 53, 55].

Several potential alternatives to raw difference computations have been proposed, such as computation of true gain [11, 12], residual gain [9], normalised gain [13, 56], average normalised gain [52], normalised change [54], ANOVA and ANCOVA on residuals or pre-post test scores [53]. Although these alternatives address the issue of measurement error, all of these methods assume that errors between participants are uncorrelated and, as such, assume that pre-test and post-test observations from one participant are independent from pre-post test observations of another participant. This assumption may not necessarily be true, as students from the same discipline, same class, and/or same university have shared variance due to the similarity of experiences, and this variance is usually overlooked [57]. One way of addressing this limitation is to analyze learning gains within a student as well as between students on a same course. Multilevel growth-curve modeling allows for estimating individual learning trajectories by fitting an overall average course curve and allowing each individual students' curve to depart from the average course curve. Moreover, using multilevel modelling it is possible to estimate what is the variance in students' initial achievements and their subsequent learning gains depending on what module they are enrolled in and whether students' initial achievements and learning gains depend on their individual differences and socio-demographic characteristics.

Several researchers have found that disciplinary differences significantly influence students' learning processes and academic performance. For example, Rienties, and Toetenel [58] found that the way teachers in their respective disciplines designed 151 modules significantly influenced how students were learning in the virtual learning environment, which in turn impacted on student satisfaction and performance. Although course characteristics are important predictors of learning, socio-demographic variables also have been found to play an important role. Thus, some researchers found that there was a gap in attainment in gender with male students being awarded higher final degree classifications than female students [59], whereas in other studies opposite was found i.e., male students were having lower initial academic achievements in comparison to female students, and the gap between males and females increased over time [60]. Ethnicity was also continuously found to be important factor in academic attainment across different levels of education, with white students having higher attainments at all levels of educational system than non-white students [61, 62]. Research also overwhelmingly shows that prior educational attainment is one of the strongest predictors of educational attainment [63, 64], with students who had high academic achievements prior to enrolling into a degree level module are more likely to have high attainments at the degree level.

In light of the challenges facing mass standardized assessments [44, 65] and assumptions on which learning gains computations are based, this study aims to test whether the estimation of a multilevel growth-curve model that accounts for the correlation of errors between participants can be effectively used in predicting students' learning gains from academic performance. As such, the first question this study will address is how much students vary in their initial achievements and their subsequent learning gains in Science and Social Science disciplines? Secondly, taking into account that previous research indicated that there are gender differences in students' achievements and progress (i.e., white students tend to perform better that students from other ethnic backgrounds), and that prior educational experience is a strong predictor of future academic success, this study will also examine whether students' initial achievements and subsequent learning gains depend of student gender, ethnicity and prior educational qualification. Finally, within learning gains research learning gains are traditionally examined in Science students and other disciplines are largely ignored. This study aims to address this gap by estimating multilevel growth-curve models separately for Science and Social Science student samples. It was hypothesized that:

H1: There will be difference in students' learning gains between Science and Social Science disciplines.

H2: There will be an effect of gender, ethnicity and prior educational qualification on students' initial achievements and subsequent learning gains.

2 Method

2.1 Setting and participants

The Open University UK is a distance-learning institution with an open-entry policy, which is the largest university in the UK. Given that, the OU is open to all people and no formal qualification requirements are present at level 1 modules. Academic performance data for 17,700 undergraduate students from Social Science and from Science faculties was retrieved from an Open University UK database. Social Science student sample comprised of 11,909 students of whom 72% were females and 28% were males with average age of M = 30.6, SD = 9.9. At the time of registering for the course 43.5% of students had A levels or equivalent qualification, 35.6% had lower that A levels, 15.7% had a HE qualification, 2.4% had postgraduate qualification, and remaining

2.8% had no formal qualification. It is important to note that in majority of UK universities A to C grades at A levels are standards for admission. The majority of students were white (86.8%) followed by black (5%), Asian (3.2%) and mixed and other (5%) ethnic backgrounds.

Science student sample comprised of 5,791 students of whom 58.2% were females and 41.8% were males with average age of M = 29.8, SD = 9.6. At the time of registering for the course 43.7% of students had A levels or equivalent qualification, 28.8% had lower that A levels, 21.6% had HE qualification, 3.9% had postgraduate qualification, and remaining 1.9% had no formal qualification. Majority of students were white (87.7%) followed by Asian (4.4%), black (3.3%) and mixed or other (4.7%) ethnic backgrounds.

2.2 Measures and Procedure

Ethics was obtained from Open University Human Research Ethics Committee (AMS ref 215140). Academic performance on Tutor Marked Assessments (TMA) was retrieved from the university database for all students enrolled to all modules within Social Science and Science faculties. TMAs usually comprise of tests, essays, reports, portfolios, workbooks, but do not include final examination scores. TMA was suitable for this study as all 111 modules used in the analysis had a minimum of two TMAs and maximum of seven TMAs. TMA grades provided enough longitudinal data for estimating students' learning gains for a period of one semester (i.e., 40 weeks). Academic performance on each TMA for each module was obtained for two cohorts of students who studied in 2013/14 and 2014/15 academic years. In total, TMA results were recorded for 111 modules across two faculties. In case of some missing TMA scores, a multilevel modelling makes automatic adjustments and estimated growth-curves on existing TMA scores and as such, some missing data is acceptable [66, 67].

2.3 Data analysis

The data was analyzed using a three-level linear growth-curve model estimated in MLWiN software [66, 67]. Identical models were estimated for Social Science modules and Science modules. In the multilevel model, level 1 variable was students' module TMA (repeated measures time variable), level 2 variable was student/participant and level 3 variable was the respective module students were enrolled in. The dependent variable was students' academic performance on each of the TMAs, with the possible minimum score of 0 and possible maximum score of 100. In line with Rasbash and colleagues [66, 67], students' academic performance was centered to the average of the course academic performance, and the time of assessment was centered to the first assessment in order to make intercept and other data parameters more interpretable. The 3-level nested structure is presented in Figure 1.

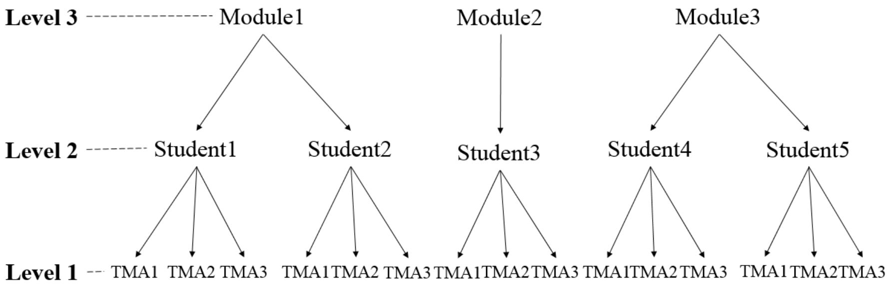

Fig. 1. A three-level data structure with repeated TMA scores at level 1

3 Results

Fitting a multilevel growth curve model to the data as opposite to single-level model (multiple linear regression) significantly improved the fit of the model for both Social Science and Science modules with the likelihood ratio test for Social Science LR = 21929.81, $p<0.001$, and for Science students LR = 11537.37, $p<0.001$ being significant. Social Science students' academic achievements were on average M = 67.6; SD = 13.7. The results of the growth-curve model estimation showed that module accounted for 6.4% of variance in students' initial achievements, and 18.5% of variance in subsequent learning gains. The student-level intercept-slope correlation was $r= 0.138$, which indicated that students with initial high achievements and students with initial low achievements progressed at a relatively similar rate. However, a module-level intercept-slope correlation indicated that students in modules with initial low achievements had much higher learning gains than students in modules with initial high achievements ($r= -.68$). Variance partition coefficient (VPC) showed that in total 3.8% of variance in Social Science students' learning gains could be attributed to the difference between modules, 56% of variance in learning gains could be attributed to individual differences, and 40% of variance was over TMAs within a student i.e., differences between assessments within the module accounted for 40% of total variance. Figure 2 represents students' actual performance, predicted growth-curves for each student, and predicted module growth curves for Social Science.

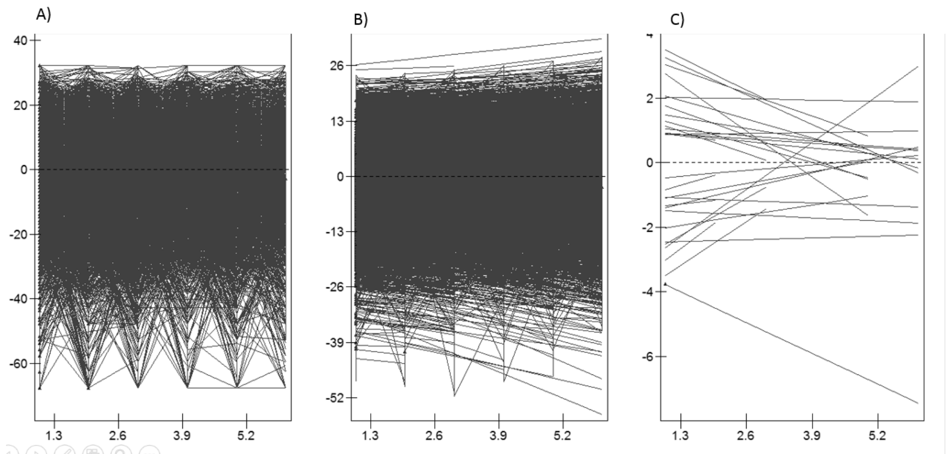

Fig. 2. A) Trellis plot for student performance on each TMA across all Social Science modules, B) Predicted student growth-curves across all Social Science modules, and C) Predicted module growth-curve for each module within Social Science.

Science students' academic achievement was on average M = 65.9; SD = 22.2. The results of the growth curve model estimation showed that 'module' accounted for 33.3% of variance in initial achievements, and 26.4% or variance in subsequent learning gains. The student-level intercept-slope correlation was r = -0.66 indicating that students with initial low achievements showed high learning gains in comparison to students with the high initial achievements. With regards to the module-level intercept-slope correlations, the correlation was r = -0.58 indicating that students in modules with initial low achievements showed higher learning gains than students in modules with initial high achievements. VPC showed that in total 26% of variance in Science students' learning gains could be attributed to the difference between modules, 52% of variance in learning gains could be attributed to individual differences, and only 22% or variance was over TMAs within a student, i.e., differences between assessments within the module accounted for 22% of total variance. Figure 3 represents students' actual performance, predicted growth-curves for each student, and predicted module growth curves for Science.

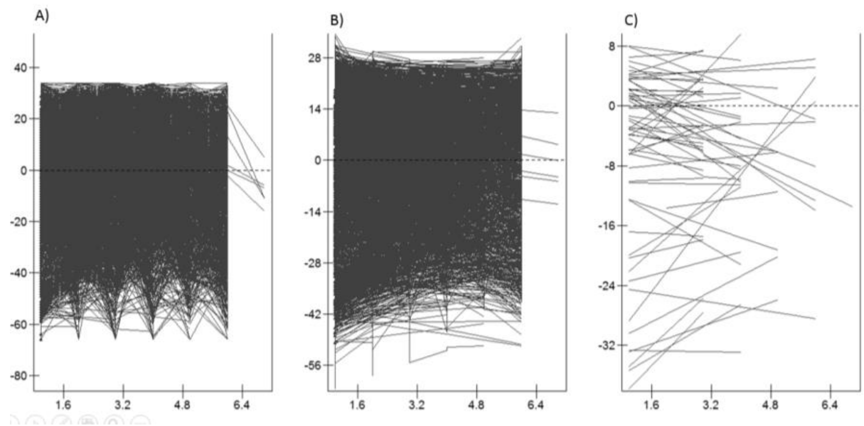

Fig. 3. A) Trellis plot for student performance on each TMA across all Science modules, B) Predicted student growth-curves across all Science modules, and C) Predicted module growth-curve for each module within Science.

Comparing Figures 2 and 3 it is noticeable that in Social Science there was a fanning out in students' predicted growth curves, which indicated that over a period of 40 weeks students with initial high achievements showed an increase in their subsequent achievements, while students with initial low achievements showed a drop in their subsequent achievements. In contrast, this phenomenon was not present for Science students, where students with initial high achievements had lower subsequent achievements, while students who initially had low achievements gradually obtained better grades. On the module level, Social Science modules showed strong fanning in, whereas it was less noticeable in Science modules. This indicated that Social Science students varied much stronger in their assessment results than Science students.

3.1 Influence of socio-demographics on learning gains

The addition of socio-demographic variables (student level predictors) further improved the fit of the model. Gender explained an additional 3% in Social Science students' initial achievements, with male students showing significantly higher learning gains than female students (Beta = 0.636, $p<0.01$), and most of this variance was due to the females having lower initial achievements, while there was no gender difference in learning gains for Science students. With regards to ethnicity, white Social Science students showed significantly higher learning gains compared to all other ethnic groups, with the biggest difference being between white and black students (Beta = -7.99, $p<0.01$), followed by the difference between white and other minority ethnic groups (Beta = -6.68, $p<0.01$), and between white and Asian students (Beta = -4.66, $p<0.01$). Overall, ethnicity accounted for an additional 3.4% of variance in Social Science students' subsequent learning gains. White Science students also showed significantly higher learning gains but only in comparison to black students (Beta = -13.07, $p<0.01$)

and Asian students (Beta = -7.31, $p<0.01$). There were no differences between white and other ethnic groups in their learning gains, and ethnicity only accounted for an additional 2.2% in Science students' subsequent learning gains.

Prior educational qualifications also explained an additional 3% of variance in both Social Science students' learning gains and Science students' learning gains. As one would expect, in Social Science students who started their course having previously obtained a postgraduate qualification showed significantly higher progress than students who only had A levels (Beta = 2.62, $p<0.05$). Students who had lower than A levels achievements or no formal qualification showed significantly lower learning gains than students who had A levels (Beta = -3.11, $p<0.01$; Beta = -6.93, $p<0.01$ respectively). There were no differences in learning gains between students who had A levels and those who already had an HE qualification. In Science, students who had HE qualification or postgraduate qualification showed significantly higher learning gains than students who only had A levels (Beta = 2.64, $p<0.01$; Beta = 8.19, $p< 0.01$ respectively). Students who had lower than A level qualification or no qualification showed significantly lower learning gains than students who had A levels (Beta = -4.59, $p<0.01$; Beta = -9.48, $p< 0.01$ respectively).

Our results overall supported our two research hypotheses and three-level growth curve models fitted longitudinal assessment data better than single-level models for both Social Science and Science disciplines. In addition, our models explained a significant portion of variance in students' initial achievements and subsequent learning gains. There were also substantial differences between Social Science and Science students in how much variance initial models accounted for, and how much additional variance socio-demographic variables accounted for in students' learning gains.

4 Discussion

The first aim of this research was to examine whether three-level growth-curve modelling on assessment scores was a better alternative to single-level models in capturing students' learning gains from module assessment data. The second aim of this project was to examine whether socio-demographic factors had any effect on students' initial achievements and subsequent learning gins. The third aim was to examine whether there was difference between multilevel models estimated for Social Science students and Science students.

The results overwhelmingly supported the superiority of multilevel growth-curve model for estimating students' learning gains in both Social Science and Science. Overall, the three-level growth-curve model for Science students accounted for more variance in learning gains than the identical model for Social Science. As such, the basic model explained variance in Science students' learning gains better than it did for Social Science students. Despite these differences, multilevel modelling was superior to single level models and as such, a more accurate method for estimating students' learning progress. The advantage of multilevel models is in that simple models are not able to detect differences between modules when looking at discipline level performance, whereas multilevel modelling accounts for those differences. This has important implications for assessing students' learning gains on an institutional level. Furthermore, this provides important policy implications when comparing learning gains across different institutions and faculties, as is currently the intention by the Teaching Excellence Framework in the UK and policy initiatives of standardized testing in the US.

In particular, an interesting finding was that Social Science students tended not to differ in progress they made regardless of their initial achievements , while amongst Science students initially low achievers were able to obtain higher learning gains over time. This finding could be due to variety of factors, but one possible explanation is that Science students' performance is more stable throughout a semester than Social Science students' performance. This is partly due to the assessments used in different disciplines to test students' knowledge. In Science, knowledge tends to be assessed using tests, workbooks and examinations, whereas in Social Science assessments are much more diverse, including essays, reports, portfolios and reviews. VPCs for each discipline further supported this interpretation, with results showing that while in Science 22% of variance in performance was across different TMAs, for Social Science students this variance was almost doubled reaching 40%. Thus, in case of Social Sciences, students may take longer to learn how to present/show their understanding and knowledge i.e., the ability to write a good essay for the first assessment does not guarantee that a student will be able to write a good report or review article for the next assessment, and hence there is greater variability in TMA scores. As such, it may be harder for Social Science students with low initial achievements to show learning gains that are higher than those with initial high achievements, despite the fact that low achievers have more room for improvement than high achievers due to obvious ceiling effects. Different patterns were observed amongst Science students, where initial low achievers outperformed initial high achievers on their rate of progress i.e., learning gains indicating that there could be a potential ceiling effect.

Another important finding of this study is that there were several modules where no learning gains or even negative ones were observed, while several modules did show large positive learning gains. Negative learning gains were previously reported in the literature and were only observed when they were estimated on students' assessments [49, 50]. However, negative learning gains that were mainly observed amongst students and modules with high initial achievements does not automatically imply that students are losing knowledge or understanding per se. However, it does highlight 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, consistency of assessment rubrics over time, and learning design [58].

Overall, our multi-level growth methodology proposed in this article starts to assess the 'distance travelled' by a student in terms of their learning over the duration of a course of study. The assessment scores that could be derived from these CBAs could facilitate a deeper dive into a more exact description of student achievement and learning. In other words, it opens the possibility of retrieving information automatically about which questions are answered well, and which skills and competencies the students are acquiring. Well contrasted CBAs can support the measurement of learning gains at the level of competences which are of interest to academics and employers alike.

With regards to the effects of demographic factors (i.e., gender, ethnicity, and prior academic qualifications) on learning gains accounted for additional 6.4% in Social Science students' learning gains, which is larger than additional variance accounted for in

Science students' learning gains (5%). Out of all socio-demographic variables, the strongest predictor for learning gains was prior educational experience/qualification with students who had A levels and above showing significantly higher learning gains than those who had below A levels or no qualification. This was closely followed by ethnicity, with white students showing highest learning gains in comparison to other ethnic groups. These findings highlight that these differences can possibly lie in different experience of HE, where white students with minimum good grades on A levels form a majority of HE students in UK [62]. However, current government plans to increase diversity of students from different ethnic and socio-economic backgrounds calls for more research into how "non-traditional" students are progressing in HE. Attracting students who are "disadvantaged" and who may be possibly the first generation in their family to attend HE also implies that universities should be actively helping those students to develop basic study skills and assist them in learning how to study effectively at an HE level [68]. Despite the fact that the initial starting point of a student might be below average, the provision of adequate support is likely to decrease the gap in students' learning gains over time between traditional and non-traditional students.

Although the results of our study are important for understanding students' learning gains in HE, this research has number of limitations that should be taken into account when interpreting and generalizing our findings. Firstly, performance data was only collected from samples of learners who were enrolled in Science and Social Science modules at one distance learning university. Because those students only usually take one module during a semester, their pace of learning could be different compared to students who are in full-time, face-to-face education and usually take four modules in one semester. Secondly, learning gains were estimated only for a limited time of a semester of 40 weeks, and as such it is possible that learning gains and observed effects of socio-demographic factors could be different when multilevel models are estimated across semesters or years. Thus, the same student could show different patterns of progress across semesters and across years. As such, future research should aim to collect longitudinal data across several years of study for full-time and part-time modules across different institutions in order to validate the generalization of our findings. Thirdly, only academic performance of students and socio-demographic data was used to estimate and explain variance in students' learning gains. Taking into account increasing use of Virtual Learning Environment (VLE) across universities, future research should aim to collect data on students' actual learning behavior, e.g., discussion forums participation, time spent going through study materials, access to additional materials and library resources. By examining differences in patterns of VLE behavior for students who make high or low learning gains, more insight could be obtained about the underlying reasons for different learning gains between students. Finally, in this study grades from the module assessments were used to estimate learning gains and detailed examination of the nature of assessments and difficulty of assessments was not taken into account. As such, future research should look more in-depth into different assessment formats and approaches used across various modules and to control for the assessment difficulty when estimating students' learning gains. Despite these limitations, the methodology of using multi-level growth modelling to understand to what extent students are making different learning gains over time seems very appropriate. In particular, our three-level linear growth-curve model can highlight differences and inconsistencies in assessment practices within and across modules, and help teachers and senior management to ensure a consistent quality of assessment provision.

References

1. Brown, G.A., Bull, J., Pendlebury, M.: Assessing Student Learning in Higher Education. Routledge (2013).
2. Boud, D., Falchikov, N.: Aligning assessment with long-term learning. Assess. Eval. High. Educ. 31, 399–413 (2006).
3. Nicol, D.: From monologue to dialogue: improving written feedback processes in mass higher education. Assess. Eval. High. Educ. 35, 501–517 (2010).
4. Tempelaar, D.T., Rienties, B., Giesbers, B.: In search for the most informative data for feedback generation: Learning analytics in a data-rich context. Comput. Hum. Behav. 47, 157–167 (2015).
5. Whitelock, D., Richardson, J., Field, D., Van Labeke, N., Pulman, S.: Designing and Testing Visual representations of Draft Essays for Higher Education Students. Presented at the Paper presented at the Learning Analytics Knoweldge conference , Indianapolis (2014).
6. Segers, M., Dochy, F., Cascallar, E. eds: Optimising New Modes of Assessment: In Search of Qualities and Standards. Kluwer Academic Publishers, Dordrecht (2003).
7. Hattie, J.: Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement. Routledge (2008).
8. Lehmann, T., Hähnlein, I., Ifenthaler, D.: Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. Comput. Hum. Behav. 32, 313–323 (2014).
9. Cronbach, L.J., Furby, L.: How we should measure “change”: Or should we? Psychol. Bull. 74, 68–80 (1970).
10. Linn, R.L., Slinde, J.A.: The Determination of the Significance of Change between Pre- and Posttesting Periods. Rev. Educ. Res. 47, 121–150 (1977).
11. Lord, F.M.: The Measurement of Growth. ETS Res. Bull. Ser. 1956, i-22 (1956).
12. Lord, F.M.: Further problems in the measurement of growth. Educ. Psychol. Meas. 18, 437–451 (1958).
13. Hake, R.R.: Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. Am. J. Phys. 66, 64–74 (1998).
14. Beck, C.W., Blumer, L.S.: Inquiry-based ecology laboratory courses improve student confidence and scientific reasoning skills. Ecosphere. 3, UNSP 112 (2012).
15. Coil, D., Wenderoth, M.P., Cunningham, M., Dirks, C.: Teaching the Process of Science: Faculty Perceptions and an Effective Methodology. Cbe-Life Sci. Educ. 9, 524–535 (2010).

16. Mortensen, C.J., Nicholson, A.M.: The flipped classroom stimulates greater learning and is a modern 21st century approach to teaching today's undergraduates. J. Anim. Sci. 93, 3722–3731 (2015).
17. Gill, T.G., Mullarkey, M.T.: Taking a Case Method Capstone Course Online: A Comparative Case Study. J. Inf. Technol. Educ.-Res. 14, 189–218 (2015).
18. Mayhew, M.J.: A Multilevel Examination of the Influence of Institutional Type on the Moral Reasoning Development of First-Year Students. J. High. Educ. 83, 367–388 (2012).
19. Riutta, S., Teodorescu, D.: Leadership Development on a Diverse Campus. J. Coll. Stud. Dev. 55, 830–836 (2014).
20. Pascarella, E.T., Salisbury, M.H., Martin, G.L., Blaich, C.: Some Complexities in the Effects of Diversity Experiences on Orientation Toward Social/Political Activism and Political Views in the First Year of College. J. High. Educ. 83, 467–496 (2012).
21. Bowman, N.A.: The Development of Psychological Well-Being Among First-Year College Students. J. Coll. Stud. Dev. 51, 180–200 (2010).
22. Trolian, T.L., Jach, E.A., Hanson, J.M., Pascarella, E.T.: Influencing academic motivation: The effects of student-faculty interaction. J. Coll. Stud. Dev. (in press).
23. Andrews, T.M., Leonard, M.J., Colgrove, C.A., Kalinowski, S.T.: Active Learning Not Associated with Student Learning in a Random Sample of College Biology Courses. Cbe-Life Sci. Educ. 10, 394–405 (2011).
24. Cahill, M.J., Hynes, K.M., Trousil, R., Brooks, L.A., McDaniel, M.A., Repice, M., Zhao, J., Frey, R.F.: Multiyear, multi-instructor evaluation of a large-class interactive-engagement curriculum. Phys. Rev. Spec. Top.-Phys. Educ. Res. 10, 020101–020119 (2014).
25. Rogaten, J., Rienties, B., Cross, S.J., Whitelock, D., Sharpe, R., Lygo-Baker, S., Littlejohn, A.: Reviewing the concept of learning gains in higher education: An affective, behaviour and cognitive perspective. Educ. Res. Rev. (Submitted).
26. Anderson, A.: An Assessment of the Perception of Learning Gains of Freshmen Students in an Introductory Course in Nutrition and Food Science. J. Food Sci. Educ. 5, 25–30 (2006).
27. Ojennus, D.D.: Assessment of Learning Gains in a Flipped Biochemistry Classroom. Biochem. Mol. Biol. Educ. 44, 20–27 (2016).
28. Hodgson, Y., Varsavsky, C., Matthews, K.E.: Assessment and teaching of science skills: whole of programme perceptions of graduating students. Assess. Eval. High. Educ. 39, 515–530 (2014).
29. Varsavsky, C., Matthews, K.E., Hodgson, Y.: Perceptions of Science Graduating Students on their Learning Gains. Int. J. Sci. Educ. 36, 929–951 (2014).
30. Douglass, J.A., Thomson, G., Zhao, C.-M.: The learning outcomes race: the value of self-reported gains in large research universities. High. Educ. 64, 317–335 (2012).
31. Richardson, J.T.E.: The attainment of White and ethnic minority students in distance education. Assess. Eval. High. Educ. 37, 393–408 (2012).
32. Farrington, D.P.: Longitudinal Research Strategies: Advantages, Problems, and Prospects. J. Am. Acad. Child Adolesc. Psychiatry. 30, 369–374 (1991).

33. Anaya, G.: College impact on student learning: Comparing the Use of Self-Reported Gains, Standardized Test Scores, and College Grades. Res. High. Educ. 40, 499–526 (1999).
34. Bowman, N.A.: Can 1st-Year College Students Accurately Report Their Learning and Development? Am. Educ. Res. J. 47, 466–496 (2010).
35. Gonyea, R.M.: Self-reported data in institutional research: Review and recommendations. New Dir. Institutional Res. 2005, 73–89 (2005).
36. Biggs, J., Tang, C.: Teaching For Quality Learning At University. McGraw-Hill Education (UK) (2011).
37. Carless, D., Salter, D., Yang, M., Lam, J.: Developing sustainable feedback practices. Stud. High. Educ. 36, 395–407 (2011).
38. Carless, D.: Trust, distrust and their impact on assessment reform. Assess. Eval. High. Educ. 34, 79–89 (2009).
39. Ras, E., Whitelock, D., Kalz, M.: The Promise and Potential of E-Assessment for Learning. In: Reimann, P., Bull, S., Kickmeier-Rust, M.D., Vatrapu, R., and Wasson, B. (eds.) Measuring and Visualising Learning in the Information-Rich Classroom. pp. 21–40. Routledge, London (2016).
40. Terzis, V., Economides, A.A.: The acceptance and use of computer based assessment. Comput. Educ. 56, 1032–1044 (2011).
41. Koedinger, K.R., Booth, J.L., Klahr, D.: Instructional Complexity and the Science to Constrain It. Science. 342, 935–937 (2013).
42. Scherer, R., Meßinger-Koppelt, J., Tiemann, R.: Developing a computer-based assessment of complex problem solving in Chemistry. Int. J. STEM Educ. 1, 2 (2014).
43. Whitelock, D.: Maximising Student Success with Automatic Formative Feedback for Both Teachers and Students. In: Ras, E. and Brinke, D.J. (eds.) Computer Assisted Assessment. Research into E-Assessment. pp. 142–148. Springer International Publishing (2015).
44. Price, M., O'Donovan, B., Rust, C., Carroll, J.: Assessment standards: a manifesto for change. Brookes EJournal Learn. Teach. 2, 1–2 (2008).
45. van Zundert, M., Sluijsmans, D., van Merriënboer, J.: Effective peer assessment processes: Research findings and future directions. Learn. Instr. 20, 270–279 (2010).
46. Whitelock, D., Gilbert, L., Gale, V.: e-Assessment tales: what types of literature are informing day-to-day practice? Int. J. E-Assess. 3, (2013).
47. Hattie, J., Yates, G.C.R.: Visible Learning and the Science of How We Learn. Routledge, London ; New York (2013).
48. Whitelock, D.: Activating Assessment for Learning: Are We on the Way with Web 2.0? In: Lee, M.J.W. and McLoughlin, C. (eds.) Web 2.0-Based E-Learning: Applying Social Informatics for Tertiary Teaching. IGI Global (2010).
49. Jensen, J.L., Kummer, T.A., Godoy, P.D.D.M.: Improvements from a Flipped Classroom May Simply Be the Fruits of Active Learning. Cbe-Life Sci. Educ. 14, 1–12 (2015).
50. Yalaki, Y.: Simple Formative Assessment, High Learning Gains in College General Chemistry. Egitim Arastirmalari-Eurasian J. Educ. Res. 10, 223–243 (2010).

51. Jensen, J.L., Kummer, T.A., Godoy, P.D.D.M.: Improvements from a Flipped Classroom May Simply Be the Fruits of Active Learning. Cbe-Life Sci. Educ. 14, (2015).
52. Bao, L.: Theoretical comparisons of average normalized gain calculations. Am. J. Phys. 74, 917–922 (2006).
53. Dimitrov, D.M., Rumrill Jr, P.D.: Pretest-posttest designs and measurement of change. Work. 20, 159–165 (2003).
54. Marx, J.D., Cummings, K.: Normalized change. Am. J. Phys. 75, 87–91 (2007).
55. Pike, G.R.: Lies, damn lies, and statistics revisited a comparison of three methods of representing change. Res. High. Educ. 33, 71–84 (1992).
56. Hake, R.R.: Lessons From the Physics-Education Reform Effort. arXiv:physics/0106087. (2001).
57. Snijders, T.A.B., Bosker, R.: Multilevel Analysis: An Introduction To Basic And Advanced Multilevel Modeling. Sage Publications Ltd, London u.a. (2011).
58. Rienties, B., Toetenel, L.: The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. Comput. Hum. Behav. 60, 333–341 (2016).
59. Mellanby, J., Martin, M., O'Doherty, J.: The "gender gap" in final examination results at Oxford University. Br. J. Psychol. 91, 377–390 (2000).
60. Conger, D., Long, M.C.: Why Are Men Falling Behind? Gender Gaps in College Performance and Persistence. Ann. Am. Acad. Pol. Soc. Sci. 627, 184–214 (2010).
61. Kao, G., Thompson, J.S.: Racial and Ethnic Stratification in Educational Achievement and Attainment. Annu. Rev. Sociol. 29, 417–442 (2003).
62. Richardson, J.T.E.: The under-attainment of ethnic minority students in UK higher education: what we know and what we don't know. J. Furth. High. Educ. 39, 278–291 (2015).
63. Plant, E.A., Ericsson, K.A., Hill, L., Asberg, K.: Why study time does not predict grade point average across college students: Implications of deliberate practice for academic performance. Contemp. Educ. Psychol. 30, 96–116 (2005).
64. Rogaten, J., Moneta, G.B.: Positive and Negative Structures and Processes Underlying Academic Performance: A Chained Mediation Model. J. Happiness Stud. 1–25 (2016).
65. Horn, C.: Standardized Assessments and the Flow of Students Into the College Admission Pool. Educ. Policy. 19, 331–348 (2005).
66. Rasbash, J., Charlton, C., Browne, W.J., Healy, M., Cameron, B.: MLwiN Version 2.02. University of Bristol, Centre for Multilevel Modelling (2005).
67. Rasbash, J., Steele, F., Browne, W.J., Goldstein, H.: A User's Guide to MLwiN. Cent. Multilevel Model. Univ. Bristol. 296 (2009).
68. Sanders, J., Rose-Adams, J.: Black and minority ethnic student attainment: A survey of research and exploration of the importance of teacher and student expectations. Widening Particip. Lifelong Learn. 16, 5–27 (2014).