Association for the Assessment of Learning in Higher Education

AALHE

PROMOTING ASSESSMENT FOR LEARNING

INTERSECTION

A journal at the intersection of assessment and learning

Summer 2018

Contents

Note from The Edition Editor 2

Using Longitudinal Cohort Datasets 3

Patterns in student ratings of instruction 6

Qualitative Analysis in Biochemistry 11

Assessing Credit Momentum and Course Enrollment 16

Comparing Assessment Data to Course Grades 23

Teachers' Dispositions in Action and edTPA 29

Note from the Edition Editor

By David Eubanks

Welcome to the 2018 Summer edition of AALHE's quarterly publication. In the call for papers, we asked you for articles on "found data" as an invitation to consider broadly the types of information that can tell us about student learning and success. The response was overwhelming, and we received enough good submissions for not only summer but also the fall edition. We selected the most quantitative articles for the edition you are reading, including both large- and small-scale studies. The topics are diverse, including a study of course evaluations, a qualitative analysis of exam questions, teacher program testing, and comparing grades to topical assessments. Thanks to a hard-working editorial staff, we are able to bring you six pieces that we hope will spark an idea or two about your own practices.

Because the *Intersection* is foremost a publication about the practice of assessment, we aim to preserve authentic author voices. As such, we have never forced articles into a standard outline format. It was tempting, given the research topics in this edition, to adopt a standardized presentation like an APA style, but we did not do that.

I hope you will find something here that can contribute to your own practice of assessment, perhaps by going through file folders and databases to look for your own "found data."

Using Longitudinal Cohort Datasets

By Michael Ben-Avie, Ph.D. and Brian Darrow, Jr.

Abstract

By incorporating all the known "bits and pieces" of data on college students at the university, longitudinal analyses driven by a research-based theoretical framework linking learning and development led to actionable results. Since 2007, the results of all of the university's assessment activities have been incorporated into these longitudinal cohort datasets with the aim of creating a complete picture of students' learning and development over time. Through merging all the seemingly unconnected data points into a cohort dataset, unforeseen patterns and anomalies emerged.

Creating Longitudinal Cohort Datasets

At Southern Connecticut State University, the Office of Assessment and Planning (OAP) conducts longitudinal cohort studies in order to identify patterns in student persistence and graduation. The aim is to alleviate the conditions that lead to student withdrawal and strengthen the conditions that promote students' academic achievement and development. There are now ten cohort datasets. As each incoming class enters the university, a cohort dataset is established (for example, 1,367 in fall 2017).

One intent of the longitudinal cohort studies was to not repeat mistakes made in the past, looking at individual pieces of data in isolation from one another. Prior to launching the longitudinal cohort studies, unconnected data were stored in different locations and in different file types. For example, students' demographic information and achievement data (GPA, grades in courses, English and Math placements) were stored in Excel workbook files by another office on campus. Within the OAP, survey data were stored in SPSS, performance-based data were stored in the online assessment system (Watermark), and scores on standardized tests (e.g., the College Learning Assessment) were stored in Excel workbook files. The disconnected nature of the data made it difficult to see patterns and anomalies across all of the diverse sources of information. Now all of this information is incorporated into the longitudinal cohort datasets, allowing the OAP to conduct more comprehensive and integrated analyses.

All undergraduate students are included in longitudinal cohort studies. The students are followed from New Student Orientation through graduation from the university, or subsequent enrollment in other colleges and universities. A cohort dataset initially contains demographic information from Banner (the university's enterprise data system) such as high school rank, high school GPA, SAT scores, gender, ethnicity, residential status, registration with Veterans Services and the Disability Resource Center, and English and Math placements. Each year, new data are added, including earned credits, cumulative GPA, registration status, and scores on surveys and direct assessments. Student IDs are used to link their demographic characteristics with their scores from surveys and assessments to create comprehensive cohort datasets.

Predictive models for student success

The longitudinal cohort datasets are used to empirically evaluate a model in which students' demonstration of the competencies that employers desire in new hires is a function of their developmental trajectories, academic habits of mind, content knowledge, interpersonal relationships, and an orientation to the future that informs goal setting and taking actions in the here-and-now to achieve desired futures. For example, data about students' academic achievement were merged with the results from the university's *Academic Habits of Mind and College Success Inventory* (AHM-CS) in order to investigate long-term student outcomes such as retention. The one-year retention rate has been steadily increasing over the last ten years to 78%. Students' cumulative GPA and the level of their future orientation were the important independent variables in a model used for predicting the probability of students staying at the university. The overall Chi-square for the retention model was 94.612 with a *p*-value < 0.001, meaning that the model was statistically significant for predicting the probability of students staying. The Nagelkerke R^2 value was 0.406. For students who stayed, the model was 93% accurate (that is, of the students whom the model predicted would stay, 93% actually stayed). Of particular note were the following odds ratios: The odds ratio for cumulative GPA was 5.624, meaning that, holding other variables constant, the odds of a student staying at the university increases 5.624 times for a one-point increase in cumulative GPA. The odds ratio for future orientation was 4.315, meaning that holding other variables constant the odds of a student staying at the university increases 4.315 times for a one-point increase in the score in future orientation.

In further analyses of the data from the ten-year study using such platforms as SPSS and IBM's Watson Analytics, there emerged a key unexpected finding. The data for these analyses were obtained from "found data"—the National Student Clearinghouse's StudentTracker for the students in the incoming classes from 2007-2014 (10,201 students). The students with high *Academic Habits of Mind and College Success Inventory* (AHM-CS) scores were still currently enrolled at the university, transferred to a private university, or graduated from the university. The students with moderate AHM-CS scores transferred to an out-of-state public university, transferred to a peer (in-state public) university, or graduated from another university. The students with low AHM-CS scores transferred to a community college or withdrew from higher education.

The analyses also demonstrated a discontinuity in students' pre-college experiences and crystallized knowledge. In other words, a sharp veer in their developmental trajectories was observed. For example, analyses conducted in IBM Watson Analytics showed that the following metrics of crystallized learning were weak predictors of retention: High School GPA (44% predictive strength), SAT Math combined with SAT Verbal (18%), SAT Math combined with Math placement score (33%), and SAT Verbal combined with High School GPA (20%). In Watson Analytics, a predictive strength of above 90% is considered noteworthy. Predictive strength is the proportion of correct classifications and uses the same algorithm as in SPSS Modeler. Ethnicity and age were also weak predictors (17% predictive strength). In terms of predictors of college cumulative GPA at time of graduation, high school GPA was a weak predictor (31% predictive strength) and, therefore, a continuity between student achievement in high school and college was not observed.

In contrast, the experiences that students had while in college, particularly those which were amenable to change, were important. In SPSS, a multiple linear regression was calculated to predict cumulative GPA at time of graduation based on students' first-year GPAs and their overall AHM-CS mean scores. A significant regression equation was found ($F(2,9026) = 11963.167$, $p < 0.001$), with an adjusted R^2 of .726. In particular, important predictors were students' levels of self-regulation, future orientation, competency in handling cognitive complexity, and sense of belonging. The extent to which their lives were complicated (e.g., routinely taking care of an older or younger relative) was also an important predictor.

Our research identified risk factors and protective factors that impact students' developmental trajectories. There are some students whose developmental trajectories were set in motion prior to attending college. These students include those who will thrive regardless of educational settings and those who do not take advantage of the opportunities afforded by the university. The students in this category do not experience intentional changes in their predicted lifepaths as a result of attending college. By way of contrast, there are students whose lifepaths are influenced by the relationships that they form with others in the university community and by their learning and developmental experiences on campus. If we only knew their incoming demographic profiles and pre-college crystallized knowledge, we would not be able to predict their college success levels.

Taken together, these findings led our institution to focus on “that which is amenable to change” instead of students’ incoming profiles. Future orientation is amenable to change in a way that demographic characteristics and crystallized knowledge are not. Likewise, crystallized knowledge garnered prior to college is set while experiences that foster the sense of belonging and fluid intelligence can be made available to everyone. A focus on those which can be modified represents a major shift in mindset away from talking about “underprepared” students. The question now is whether the university is prepared to educate all students. Were the focus on students’ unchangeable demographics or past learning, some may have felt that students’ incoming profiles alleviated them from the responsibility of preventing student failure. To counter this, the university established, for example, a First-Year Experience program with a living-learning community component for first-generation college students in which the staff members were also the first in their families to attend college.

Conclusion

It takes a great deal of time and effort to merge “bits and pieces” of data in order to set up and maintain longitudinal, cohort datasets. Despite this challenge, our analyses have indicated the value in assessing students using a longitudinal approach. The cohort analyses of “found data” indicate that focusing on those which are amenable to change may be the most important implication of all. The research conducted by the OAP continues to influence policy decisions at the university that improve the academic experiences of our students. The unique advantage of our assessment strategies is that there is no need to scramble when the university wants to make a decision that is informed by data.

Dr. Michael Ben-Avie was recently the Director of the Office of Assessment and Planning at Southern Connecticut State University and is now Senior Director of Learning Assessment and Research at Quinnipiac University. He can be contacted at michael.ben-avie@quinnipiac.edu

Mr. Brian Darrow, Jr. is the Research Scientist for Assessment and Planning and an Adjunct Professor of Mathematics at Southern Connecticut State University and can be contacted at DarrowB1@SouthernCT.edu.

Patterns in Student Ratings of Instruction

By Lisa K. Bonneau, Ph.D.

Abstract

It is sometimes hypothesized that the grade a student earns in a course may affect the student's appraisal of the course instructor's teaching performance. To investigate this idea empirically, we conducted an analysis to explore the relationship between course grading outcomes and IDEA course evaluation outcomes in sections of 100-level courses offered in a single academic year. This analysis observed only a weak relationship between these outcomes and were similar to results of other published studies. However, interesting patterns were noted among colleges, schools, and delivery methods that are worthy of future investigation when a larger data set can be used.

Introduction

Many institutions use student ratings of instruction to improve teaching and evaluate faculty. It is natural, given the wealth of course section data available within an institution, to ask whether there are other ways to utilize the data to impact student learning. This paper analyzes some connections between course evaluations and course grades.

Some faculty believe that the grade a student earns in a course may affect that student's appraisal of teaching performance (Benton and Ryalls, 2016), but the extant research findings are mixed. In a study of 50,000 courses, Centra (2003) found a minimal effect of grades on evaluations, but Kockelman (2001) found a positive correlation between average course grades and student ratings in more than 2,500 engineering courses. Griffin (2004) also showed a positive relationship between grading leniency and evaluation scores.

Method

The University of South Dakota uses IDEA Learning Essential ratings as a course evaluation tool. For this study, IDEA scores were analyzed to determine the likelihood that course-level student grades impact the faculty rating and to determine patterns in instructional scores across colleges and delivery methods at the institution.

Since IDEA does not provide course evaluation data at the student level, the course section was treated as the primary unit of analysis. The IDEA Learning Essentials rating form asks students to rank their progress on thirteen learning objectives and provide scores for Excellent Teacher and Excellent Course based on a 1-5 Likert-type scale. For each course section included in the analysis, several outcome variables were used or calculated: the percentage of students earning a C or higher in the section; the percentage of students earning an A in the section; the number of students enrolled in the section; the adjusted converted average IDEA scores for Excellent Teacher, Excellent Course, and Summary Evaluation (an indicator which describes progress on relevant objectives); and the raw IDEA scores for Excellent Teacher, Excellent Course, and Summary Evaluation.

The adjusted converted average IDEA score is derived from the following variables: student background preparation, work habits, motivation, and class size. Because the average student rating of progress is different for each of the learning objectives, these averages are rescaled to convert averages to 50 and standard deviations to 10. The scores are weighted by the faculty member's rating of the importance (relevance) of each objective. The raw score is calculated from the students' rank of their progress on objectives, plus their ratings for Excellent Teacher and Excellent Course (Hoyt and Lee, 2002).

A total of 581 100-level course sections from two semesters were analyzed. Course sections that were individual in nature (independent studies, experiential learning, lessons, etc.) were not included, and sections with fewer than 10 students enrolled were also eliminated from the analysis. Table 1 displays summary statistics for the variables for both the adjusted scores and the raw scores.

Variable	n	Mean	SD	Min	Max
Excellent Teacher	581	48.4/49.2	9.9/9.68	-1.4/5.3	68.9/62.1
Excellent Course	581	47.2/48.5	10.0/9.68	-6.1/5.3	70.9/65.3
Summary Evaluation	581	48.1/48.1	8.9/8.9	1.0/1.0	68.0/68.0
Percent C or Higher	581	88.3	11.8	35.7	100.0
Percent A	581	46.6	27.7	0.0	100.0
Enrollment	581	32.6	30.87	10.0	261.0

Table 1. Descriptive Statistics for Variables of Interest for Adjusted and Raw IDEA Scores and Calculated Variables

The university expects faculty to maintain an average IDEA score of 45 or higher for the Excellent Teacher, Excellent Course, and Summary Evaluation. For the two semesters analyzed, 70% of sections had an IDEA score at 45 or higher for the Summary Evaluation and the Excellent Teacher variables while 64% of sections had a value of 45 or higher for the Excellent Course variable. Pearson's *r* coefficients were used to measure correlation among IDEA score variables (both adjusted and raw) and course grade variables, and the correlation matrix is displayed in Table 2. Statistically significant correlations ($p < 0.05$ shown in bold) were found between most variable pairs though the low magnitude of these values suggests that—from a practical perspective and given the limited number of sections evaluated—these variables do not appear to be closely related, which undermines the suggestion that course grades affect an instructor's IDEA score ratings. Note that low p-values are not evidence that the correlations are different from each other, just that they are non-zero. In most cases, the raw scores have larger correlations than the adjusted scores, probably because IDEA's adjusted figures incorporate information unrelated to the identified question. The largest correlation was between the fraction of A grades and the raw scores for Excellent Course.

	Excellent Teacher	Excellent Course	Summary Evaluation
Percent C or Higher	0.05/0.15	**0.13/0.25**	0.07/0.07
Percent A	**0.10/0.22**	**0.18/0.32**	**0.12/0.12**
Enrollment	-0.03/**-0.11**	-0.04/**-0.11**	**-0.09/-0.16**

Table 2. Correlation Matrix with Pearson r Coefficients for Adjusted Scores/Raw Scores

Due to the variety of 100-level courses at the institution, data were further subdivided at the college level to help elucidate any particular patterns in correlations. Average response rates and enrollment numbers for the colleges with more than twenty sections in the analysis are listed in Table 3. Table 4 displays the correlation coefficients found between IDEA variables and grading variables for each college with twenty or more sections of 100-level courses. As with the university-wide figures reported above, most coefficients computed for individual colleges were fairly small. (Due to the section exclusions used in this analysis, limited *n* values are available for several colleges.) For example, the College of Arts and Sciences, which had the largest source of course sections, generated no coefficients with a magnitude of

0.20 or greater for adjusted scores, and when analyzing raw scores, none was greater than 0.30. The two remaining colleges with twenty or more sections produced somewhat higher *r* values. Health Sciences had the highest *r* values but the lowest number of course sections. Interestingly, the College of Fine Arts had a negative correlation between grades and IDEA scores when analyzing adjusted IDEA scores—but not for raw scores. It is also interesting to note that there was a negative correlation between IDEA scores and course enrollment for both Arts and Sciences and Health Sciences but not for Fine Arts.

	Average Response Rate	Maximum	Minimum
Arts & Sciences (*n*=454)			
Response Rate	59.2%	100%	4.55%
Enrollment		261 Students	10 Students
Fine Arts (*n*=71)			
Response Rate	56.7%	91.67%	18.18%
Enrollment		135 Students	10 Students
Health Sciences (*n*=28)			
Response Rate	55.9%	96.3%	25%
Enrollment		72 Students	16 Students

Table 3. Response Rates and Section Enrollment Values for Each College

	Excellent Teacher	Excellent Course	Summary Evaluation
Arts & Sciences (*n*=454)			
Percent C or Higher	**0.13**	**0.23**	**0.18**
Percent A	**0.21**	**0.29**	**0.23**
Enrollment	**-0.12**	**-0.12**	**-0.18**
Fine Arts (*n*=71)			
Percent C or Higher	-0.05	0.01	0.06
Percent A	-0.14	0.02	0.07
Enrollment	0.01	-0.03	0.06
Health Sciences (*n*=28)			
Percent C or Higher	**0.41**	**0.39**	0.36
Percent A	**0.45**	0.31	**0.45**
Enrollment	-0.10	-0.20	-0.16

Table 4. Correlation Matrix by College (Pearson r Coefficients, Raw IDEA Scores)

Many 100-level courses are general education courses, and as such, they are delivered both online and in traditional classrooms. Of the sections in this analysis, 151 were offered in an online format (Table 5). At this stage, sample sizes are not large enough to determine whether courses show similar patterns regardless of delivery method.

	Excellent Teacher	Excellent Course	Summary Evaluation
Traditional (*n*=430)			
Percent C or Higher	0.02/0.11	**0.11/0.23**	0.05/**0.18**
Percent A	**0.11/0.23**	**0.17/0.32**	**0.11/0.27**
Online (*n*=151)			
Percent C or Higher	0.16/**0.26**	**0.20/0.35**	0.13/**0.28**
Percent A	0.12/**0.20**	**0.24/0.32**	0.15/**0.25**

Table 5. Correlation Matrix by Course Delivery (Pearson r Coefficients, Adjusted/Raw)

Discussion
The above results mirror those of Kockelman (2001), yet the effects are weak enough that they do little to undermine Centra (2003). Though statistically significant correlation values were found between course grading outcomes and IDEA course evaluation outcomes across the colleges, those findings should be viewed with caution as data for the analysis were limited in scope and included variables from only two semesters. It remains possible that more pronounced relationships may exist in particular instructional contexts (e.g. by department, STEM courses, and/or mode of instruction). For example, courses in Arts and Sciences are largely general education in nature (with the exception of a few gateway courses in some fields) while the courses in the Health Sciences are all “for majors” courses. The School of Fine Arts is a mix of general education courses, “for majors” courses, and performance courses (band, choir, etc.). When more data become available, analysis at the department level might elucidate additional patterns within colleges. It is possible that IDEA scores in gateway courses might have a higher correlation with grades than those in general education courses.

With these provisos, we can consider possible meanings for the modest correlations. First, raw scores are more plausibly related to grades than adjusted scores, since they have not been recalculated to incorporate information from other sections. The slight negative correlations between IDEA averages and class size are not surprising, since larger classes present more challenging learning environments. The most interesting remaining pattern is the positive correlations between IDEA score averages and grades for the colleges, which is lowest for Fine Arts, and highest for Health Sciences. It is possible that when students can see clear connections among their coursework, learning, and grade, they give higher ratings to faculty. A syllabus analysis might provide evidence that the less subjective the material is, the higher the rating averages.

Conclusions
Though some modest correlations exist between IDEA outcomes and course grades, it does not appear that course grading outcomes are strongly associated with student evaluations of instructional quality on average. It can also be said that interesting patterns and associations do exist and are worthy of further investigation. For example, it is possible that student engagement may play a role in a student’s evaluation of a course. This might be examined by evaluating the patterns observed alongside the results for the IDEA questions “As a rule, I put forth more effort than other students on academic work” and “I really wanted to take this course regardless of who taught it."

Because some of the most interesting patterns emerged in data segregated by college, future work will further partition data by departments within Arts and Sciences and analyze data for general education sections and STEM gateway sections. Due to low sample sizes, such an analysis will need to span multiple academic years. In addition, the institution is currently reviewing syllabi. A study is planned to determine whether course syllabi reflect the IDEA objective questions that faculty select as important. It is possible that there is a lack of connection between what faculty select as important for their course on the IDEA form and what is outlined in the syllabus. If this is the case, the IDEA scores for the faculty member could be affected, explaining some of the noise in the data.

References

Benton, S.L., & Ryalls, K.R. (2016). Challenging misconceptions about student ratings of instruction. *IDEA Paper #58*. Retrieved from https://www.ideaedu.org/Portals/0/Uploads/Documents/Challenging_Misconceptions_About_Student_Ratings_of_Instruction.pdf

Centra, J.A. (2003). Will teachers receive higher student evaluations by giving higher grades and less course work? *Research in Higher Education*. 44(5):495-518.

Griffin, B.W. (2004). Grading leniency, grade discrepancy, and student ratings of instruction. *Contemporary Educational Psychology*. 29(4):410-425.

Hoyt, D.P., & Lee, E.J. (2002). *Technical Report No. 12: Basic Data for the Revised IDEA System*. The Individual Development and Educational Assessment Center, Manhattan, KS.

Kockelman, K.M. (2001). Student grades and course evaluations in engineering: What makes a difference. *ASEE Annual Conference Proceedings, Paper #2793*, Albuquerque, NM.

Lisa Bonneau is Director of Assessment at University of South Dakota, and can be reached at Lisa.Bonneau@usd.edu.

Qualitative Analysis in Biochemistry

By Alexandra Greb and Mona Monfared

Abstract

Answers on written exam questions in a high enrollment biochemistry course were analyzed to gain insight into which concepts posed the greatest difficulty for students. The analysis involved outlining which key concepts were being tested in a question, creating a coding scheme based on errors in student answers, and coding a number of exams to identify which errors were most common. Questions were selected for analysis by counting points awarded for each question and selecting the questions with the lowest student performance. The process of listing key concepts tested by each question allowed us to discover flaws in the grading rubric – while the rubric was technically correct, the practice of connecting questions to key concepts helped identify where points were awarded for concepts that were not crucial in the lesson and rubric parameters that were overlapping. Coding the written response answers provided quantitative data on which errors were most common. These errors can be linked to student misconceptions and are informative in guiding the way the material is presented, what practice can be given to students, and how the concepts are assessed.

Introduction

Written response questions on summative exams are rich sources of data that can provide direction for changes in instruction and assessment. We reviewed student written responses from a midterm exam at University of California, Davis in an upper division biochemistry course (BIS102 - Structure and Function of Biomolecules). Most students in BIS102 are enrolled in the College of Biological Sciences, which includes majors such as Genetics and Genomics, Microbiology, and Biochemistry and Molecular Biology. This study is based on six sections of BIS201 from the past three years, where course enrollment ranged from 202 to 435 students. The summative assessments are in-class exams that use a combination of multiple choice questions (graded by scantron) and written response questions (graded by Teaching Assistants (TAs) using a rubric created by the instructor). The written response questions typically ask the students to do one or more of the following: perform biochemical calculations, draw biochemical structures or graphs, and interpret data.

Results and Discussion

An analysis of written response data (N=202 students) from a Fall 2017 midterm exam helped to identify:

1) which questions were most difficult for students and had the greatest range of scores, 2) key concepts being tested in each question to examine alignment between assessment and intended learning objectives, and 3) common student misconceptions/errors. Written response questions accounted for 57% of the exam points and were graded by TAs using a rubric. The instructor created rubrics before the exam and refined them afterwards by reviewing 5-10 exams with a TA and making appropriate adjustments based on student responses. The TAs also conferred with the instructor during the grading process to resolve any ambiguities in grading and adjust the rubrics as necessary.

For our analysis, we first looked at the points awarded for each of the four questions and (where relevant) their sub-questions. The point averages and distributions were informative, as they did not correlate to what we assumed were the most difficult portions of the exam. Question 3, for example, involved drawing the chemical structure of a large molecule with many parts, which most students did correctly. Question 1 was asking for a drawing with fewer parts, but students performed less well on this question. Table 1 provides a listing of the written questions, organized by average percentage of points earned (from highest to lowest). The instructor labeled the question by topic assessed as well as by Bloom's taxonomy of cognitive domains using the Blooming Biology Tool as a guide (Crowe, Dirks, & Wenderoth, 2008).

	Total possible points	Average points awarded	Standard deviation	Average Percent of total points earned	Topic assessed	Bloom's Taxonomy level
Question 3b	10	9.2	1.6	92.3	Lipid structure	Remember
Question 2a	10	7.7	2.8	77.1	Protein techniques	Apply
Question 3a	3	2.2	1.1	74.6	Lipid structure	Apply
Question 4	13	9.3	3	71.6	O_2 binding proteins	Remember
Question 1	12	6.4	3.1	53.6	Protein structure	Apply
Question 2c	4	1.4	1.1	36.1	Protein techniques	Apply
Question 2b	5	1.7	1.6	34.9	Protein techniques	Analyze

Table 1. Points possible and awarded for written response questions on Midterm Exam (N = 202).

Our analysis showed that Questions 3 and 4, which measured understanding of lipid structure and oxygen binding proteins, had a higher percentage of points earned relative to the other questions. We identified Questions 1 and 2 as good targets for identifying student misconceptions because students earned lower points on these questions and the scores had relatively large standard deviations.

For the remainder of this article, we will focus on describing our process and results associated with analyzing written responses for Question 1. We first listed the key concepts assessed by the question. Our list showed that Question 1 tested the following key concepts: 1) Structure of amino acid R groups, 2) How amino acids are connected in a protein, 3) The structures of amino acids ("residues") in a protein and how they look if they are in the middle or end of a protein, and 4) How amino acids form ionic interactions (relationships between positively and negatively charged groups) in proteins. The exercise of listing the key concepts tested was illuminating, as we built grading rubrics based on what we thought were the most complete answers. We realized that part of the grading rubric included 1.5 points for putting the amino acids into the correct subunit, but that is not a key concept that is important for the course. In the future, we will change the question accordingly. Our analysis helped us align the learning objectives, written exam questions, and grading rubric.

It was a surprise to see the low scores on Question 1, on protein structure. Key concepts tested in this question are repeated throughout the quarter and outlined in lecture and online course materials. Students also received practice questions on these key concepts. Coding student errors and misconceptions was helpful for understanding the reason for the low scores. The instructor compiled a list of errors made on Question 1 by reviewing 20 randomly selected exams (Table 2). While creating this coding table, we found that the rubric was flawed in how it assessed understanding of Key Concept #4 (how amino acids form ionic interactions). The rubric underestimated students understanding of Key Concept 4—in particular, those who made Code error 1 (Table 2). The rubric was not consistent in the way students were awarded for points on that part of the question. This likely contributed to the large standard deviation observed in points earned for Question 1 (Table 1). We determined that assessing the key concepts separately would help to truly gauge student understanding and award points fairly on an exam. Had we created the key concept list before grading began, we would have caught this error. In the future, we will work with the TAs to list key concepts as we create the rubrics.

Code #	Description of error	% of students with code (N=30)
7	Draw as residues but not on same peptide	67%
1	Drawing free amino acids rather than residues in a protein	57%
3	R group structure incorrect (minor or major error)	27%
2	Ionic interactions not between + and – groups or no interactions drawn	23%
0	No error	13%
4	His location: Not drawing His at C-terminal end so His lacks COO- group	10%
8	Salt bridges not between correct groups although charges are correct	0.3%
5	Ionization of His: drawing His as a free amino acid but without COO- group	0.07%
6	No structures	0.03%
9	Asp at C-terminal end rather than middle	0.03%
10	Only R groups drawn	0.03%
11	Part of peptide bond incorrect	0.03%

Table 2. Coding scheme for Question 1 on Midterm Exam

Coding the answers for Question 1 showed that the most common error in the sample of responses reviewed was code 7, with 67% of students making this error. This error is associated with the concept mentioned earlier (putting the amino acids into the correct subunit), which we do not consider a key concept for students at this stage of their education. The finding reflects a lack of course coverage of the concept, for example through examples or practice homework. If we had created the key concept list before designing the grading rubric, we would not have awarded points for this concept in the rubric.

The high rate of error 1 was very surprising, showing that that Key Concept 2 (How amino acids are connected in a protein) was not grasped by the students despite being covered in great extent in lectures and homework. However, students had never seen the problem written in the way it was given on the test. Asking them to unpack the description of protein structure in a new context may have been the cause of the errors.

A larger point is that students typically struggle with applying their knowledge to a new context. Anecdotally, we've seen students in this course approach the content by "memorizing the steps" to solve the problems rather than attempting a conceptual understanding. An example of this is a student asking during office hours, "So can I say we always subtract x from the left side of the equals sign?" We do many different types of biochemical calculations in this course, and that kind of "pattern" finding is common but not helpful for success. The test question was designed to go beyond the algorithmic approach many students rely on, and the data show that most students did not grasp one of the most basic concepts in our course. This question can be improved by rewriting it to create a more focused context, so that the question is assessing only the intended concepts (Ahmed and Pollitt, 2007). The finding that students missed a basic concept also has implications for instruction; we need to address the key concepts better in the course, whether by providing more practice with contextual descriptions of proteins or discussing common student misconceptions. Lastly, we wondered if the issue stemmed from reading comprehension. We emphasize text reading to our students to hone reading comprehension skills, which is arguably more important than any biochemistry concept we discuss. This is complicated, however, as reading comprehension is not an explicitly being tested on this exam. We would need to either make it an objective and provide instruction and practice with feedback before the exam, or we need to be more careful about how questions are worded.

Conclusions

A strength of the exam analysis method that we used is that it allows an instructor to gather quantitative data on student performance on open ended question types, even in a high enrollment course. However, the link between student errors on exams and student misconceptions may not always be present. Student errors may arise from the way a question is worded, the way the rubric is designed by the instructor, or how the rubric is interpreted by graders. Despite the fact that the link between student error and misconception may not always be made, the analysis is useful as a reflective teaching practice. One challenge is finding the time to do this during a course. If the same course is taught multiple times in a year, the analysis from one quarter can inform teaching decisions the following quarter, but that is not always the case. The time it takes to create the coding scheme is minimal, but tabulating points earned on questions and coding a number of exams takes time and can be difficult to do while the course is underway.

In this study, the process of digging into test questions and answers shed light on flaws in the assessment process. Identifying key concepts that were intended to be tested by a question allowed us the opportunity to reflect on alignment between our rubrics, exam questions and learning objectives. Listing key concepts with TAs before they grade an exam question can ensure that every point awarded on a rubric is directly related to a learning objective. The TA graders can then be consistent in their grading and efficiently score exams, but they will often be too literal in their interpretation of the rubric and may not view the student answers holistically. Identifying the key concepts being tested in each question with the TAs before discussing the rubric allows for more consistent and accurate grading.

Looking at student responses and creating a coding scheme to organize the qualitative data helps us see patterns in student understanding and guides us in making evidence-based decisions about teaching and assessment. Whether analyzing written response questions inspire changes in exam questions or the way a concept is presented in a course, working with this data can help an instructor be more connected to what the students are experiencing in a class. This is especially helpful in a large enrollment course, where the challenge is to find ways to connect with all your students.

References

Ahmed, A. & Pollitt, A. (2007) Improving the quality of contextualized questions: an experimental investigation of focus, Assessment in Education, 14:2, 201-232, DOI: https://doi.org/10.1080/09695940701478909

Crowe, A., Dirks, C., and Wenderoth M.P. (2008). Biology in Bloom: Implementing Bloom's Taxonomy to Enhance Student Learning in Biology, *CBE Life Sciences Education,* 7(4), 368-381, DOI: https: //doi.org/ 10.1187/cbe.17-06-0109

Alexandra Greb is an undergraduate researcher at University of California-Davis, and can be reached at acgreb@ucdavis.edu.

Mona Monfared is Lecturer PSOE at University of California-Davis, and can be reached at mmmonfared@ucdavis.edu.

Assessing Credit Momentum and Course Enrollment

By Renata Opoczynski, Susan Richter, Bethan Cantwell, and Mark Largent

Abstract

Specifically, MSU's Learning Analytical Group, in coordination with Institutional Studies and the Office of the Associate Provost for Undergraduate Education, looked at course registration data and discovered that a lower percentage of first-year students had registered for 15 credits in each semester over the last decade and an even lower percentage for 30 credits over their first 12 months at MSU. We used analytics methods to learn that students who attempted fewer than 30 credits in a calendar year graduated at lower rates, and that this persisted across all racial, ethnic, and gender categories. Michigan State's "Go Green, Go 15" campaign was created to eliminate roadblocks to student success.

Introduction

Credit-momentum initiatives emerged out of efforts to understand the link between the pace at which students accumulate credits with time-to-degree and graduation rates. Adelman (2006) found that students who accrue credits at a consistently brisk pace are more likely to graduate in a timely manner. The number of full-time undergraduate students attempting 12 credits or more per semester across all of higher education has decreased in recent years (Attewell & Monaghan, 2016). Additionally, the average time-to-degree for undergraduate degrees recipients has increased, with students from the high school class of 1972 taking an average 4.69 years, compared to 4.97 years for students from the high school class of 1992 (Bound, Lovenheim, & Turner, 2010). Viewed another way, 45 percent of students graduating in 1977 had earned their degree in four years or less, while only 31 percent had done so in the 1990s (Bound, Lovenheim, & Turner, 2010).

While limited research has explored the benefits of credit momentum, research that does exist found that full-time undergraduate students who attempted 15 credits per semester in their first year had significant and positive effects in the form of higher graduation rates compared to students who initially attempted 12 credits (Attewell & Monaghan, 2016; Belfield, Jenkins, & Lahr, 2016). Attewell and Monaghan (2016) note particularly pronounced effects among Black, Hispanic, and first-generation students as well as those students with relatively less-rigorous high school backgrounds. Additionally, by analyzing the penalty of taking 12 credits on students who would otherwise take 15, they emphasized these positive effects are independent of selection effects.

This paper discusses how Michigan State University (MSU) utilized existing enrollment data to highlight university policies and procedures (both actual and assumed) that hindered student success. A credit-momentum campaign, Go Green, Go 15, was created and implemented in response to the discovery.

Found Data at MSU

MSU keeps detailed information on course enrollments, recorded through a series of snapshots taken at key points in the semester. MSU's Office of Institutional Studies delved into the University's Student Information System (SIS) course registration data to examine student-level course-enrollment patterns for each student through a series of frozen snapshots stored in the data warehouse. These snapshots showed enrollment at specific points in time across the semester allowing student tracking from summer academic orientation and the first day of class through the last official drop-point (quarter semester) to last day of class. Registration snapshots provided a picture of student course enrollment changes across the semester. Snapshot data were then combined with course outcomes (grades) to measure academic progress.

Additional data stored within the SIS warehouse included race/ethnicity, gender, FAFSA data, high school GPA, ACT/SAT scores, entering major preference, and first-generation status. Academic data from enrollment snapshots were combined with demographic and academic preparation measures to establish a comprehensive picture of each enrolled student. From this data-set, researchers were able to compare students with similar characteristics but different course enrollment behaviors in order to conduct further analysis. Each year, MSU enrolls an entering class of approximately 8,000 first time students in addition to 1,500 transfer students. For ease of analysis, students were placed into entering cohorts to follow credit accumulation by semester and academic year. The combination of the point in semester data and overall cohort grouping created a clear longitudinal picture of students' credit-taking patterns and outcomes showing overall credit momentum at the institution. Using the existing data, clear patterns of credit accumulation behavior emerged.

Needs Assessment

MSU's Learning Analytical Group, in coordination with Institutional Studies and the Office of the Associate Provost for Undergraduate Education, looked at course registration data and discovered that the percentage of incoming MSU first-time-in-any-college (FTIAC) undergraduate student who took 15 or more credits per semester had steadily declined over the last decade, from a high of 43 percent in 2006 to a low of 28 percent by 2016. Similarly, the proportion of students who attempted 30 or more credits in one (calendar) year had declined from a high of 44 percent in 2006 to a low of 34 percent in 2015.

Snapshot data also provided the ability to understand the level of credit melt, which is the difference between credits enrolled in on the first day of the term and attempted credits as of the end of the term. The average credit melt for fall is one-half a credit, and for spring, it is one credit. Understanding the amount of credit melt that occurs during the semester is essential for ensuring that a student will attempt at least 15 credits by the end of the semester. For example, if students enroll in an average of 15 credits on the first day, the average credit melt could result in attempted credits between 13 and 14 credits.

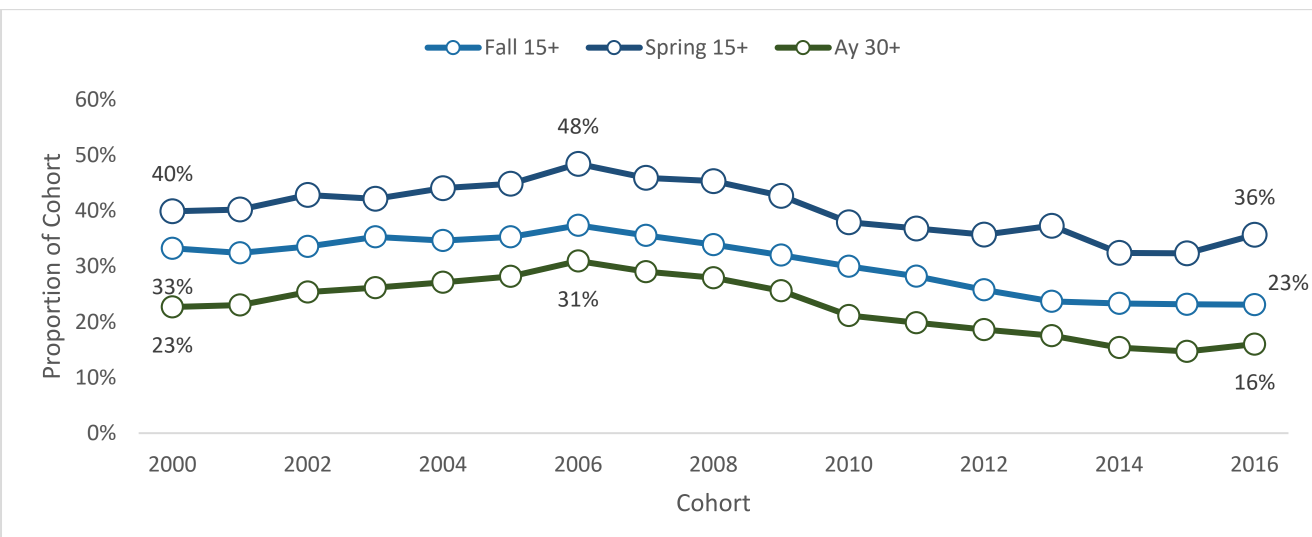

Figure 1: Proportion of students that attempt 15+ credits in Fall or Spring or 30+ credits in an academic year

The initial analysis indicated that it was necessary to take a closer look at the correlation between the number of credits attempted in the academic year and persistence and graduation rates. We found that overall first returning fall persistence (the percent of students that returned for their second fall semester) was nine percentage points less for students who did not attempt at least 30 credits in their first academic year compared to the overall population**.** Second returning fall persistence (the percent of students who returned for their third fall semester) decreased to 83 percent for students who did not attempt at least 30 credits in their first academic year. Additionally, we found positive correlations between credit momentum and cumulative GPA. Students who did not attempt 30 credits in an academic year had lower cumulative GPA and higher rates of academic probation. Finally, six-year graduation rates were seven percentage points lower for students who did not attempt at least 15 credits in either fall or spring semesters compared to students who attempted 30 or more credits in their first academic year. The median credit load for students, who did not attempt at least 15 credits in the fall or the spring semesters, was 26 credits in an academic year. The 25th percentile for this group was 25 credits, while the 75th percentile was 28 credits. Therefore, almost half the students in this range were only three to four credits away from attempting 30 credits.

To assess the impact of credit momentum, propensity score matching, as described in Attewell and Monaghan (2016), was employed to minimize selection bias and present an unbiased estimate of the relationship between credit momentum and student success outcomes. Students in the low-credit group did not attempt 30 credits in a calendar year and were compared to the matched high-credit group to estimate the penalty on graduation rates associated with not taking 30 credits. The graduation rate penalty for taking fewer than 30 credits is statistically significant (p-value < 0.05), with students in the lower credit category having a six-year graduation rate 13 percentage points lower than the 30-credits group.

There are two components to credit momentum in our model: the amount of credit melt that occurs over the semester and the number of attempted credits in a calendar year. This analysis guided a campus-wide campaign that emphasized the value of attempting 30 credits in a year and for students to work with an academic advisor to develop a plan for completing their degrees.

First-year outcomes

As a result of the Go Green, Go 15 campaign, 51 percent more first-year students attempted 15 credits in their first fall semester from 2016 to 2017. There was also a 19 percent increase for spring semester and 63 percent increase in the proportion of students attempting 30 credits in their first twelve months at MSU. Fall and spring semester outcomes have shown no significant decline in students in good academic standing and in cumulative GPA for students attempting 15 or more credits a semester despite the fact that a significantly larger number of students were attempting 15 or more credits.

The campaign uncovered another roadblock to student success. Many of the courses at MSU overlapped with one another and were intensely scheduled during certain days and time slots making it difficult for many students to enroll in 15 credits or courses they needed in their first semesters at MSU. To address this concern, an effort was put forth to distribute classes more broadly throughout the day and week, improving students' ability to register for a larger number of classes each semester.

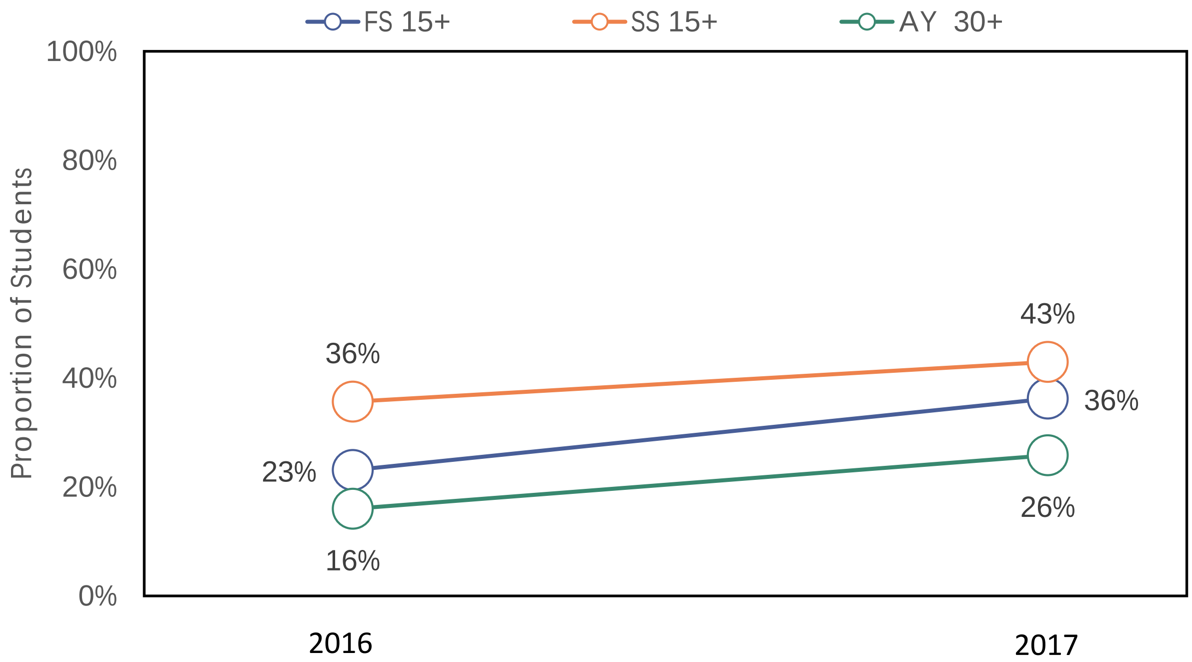

Figure 2: Change between 2016 and 2017 in the proportion of students attempting 15+ credits in Fall or Spring or 30+ credits in an academic year.

For students at MSU who were able to attempt 30 or more credits in their first year at MSU, the benefits of the higher level of credit momentum were evident. During the 2017-18 academic year, about 26 percent of MSU students attempted 30 or more credits in the academic year, and of these, 97 percent finished their first year in good academic standing (defined at MSU as having a cumulative GPA of 2.0 or greater on a 4.0 scale). Comparatively, among first-year students at MSU in the 2017-18 academic year who attempted fewer than 30 credits, 88 percent finished their first year in good academic standing, a difference of nine percentage points.

Similar positive benefits of higher-credit momentum were evident for the MSU students who attempted 15 or more credits in either the fall or the spring semesters of 2017-18. Thirty-seven percent attempted 15 or more credits in the fall, and approximately 95 percent were in good academic standing by the end of the year, compared with 90 percent of students who attempted fewer than 15 credits. Likewise, in the spring semester, when nearly 43 percent of students attempted at least 15 credits, 97 percent of the students at 15 or more credits were in good academic standing. However, for the students who attempted fewer than 15 credits in the spring semester, only 6 percent were in good academic standing.

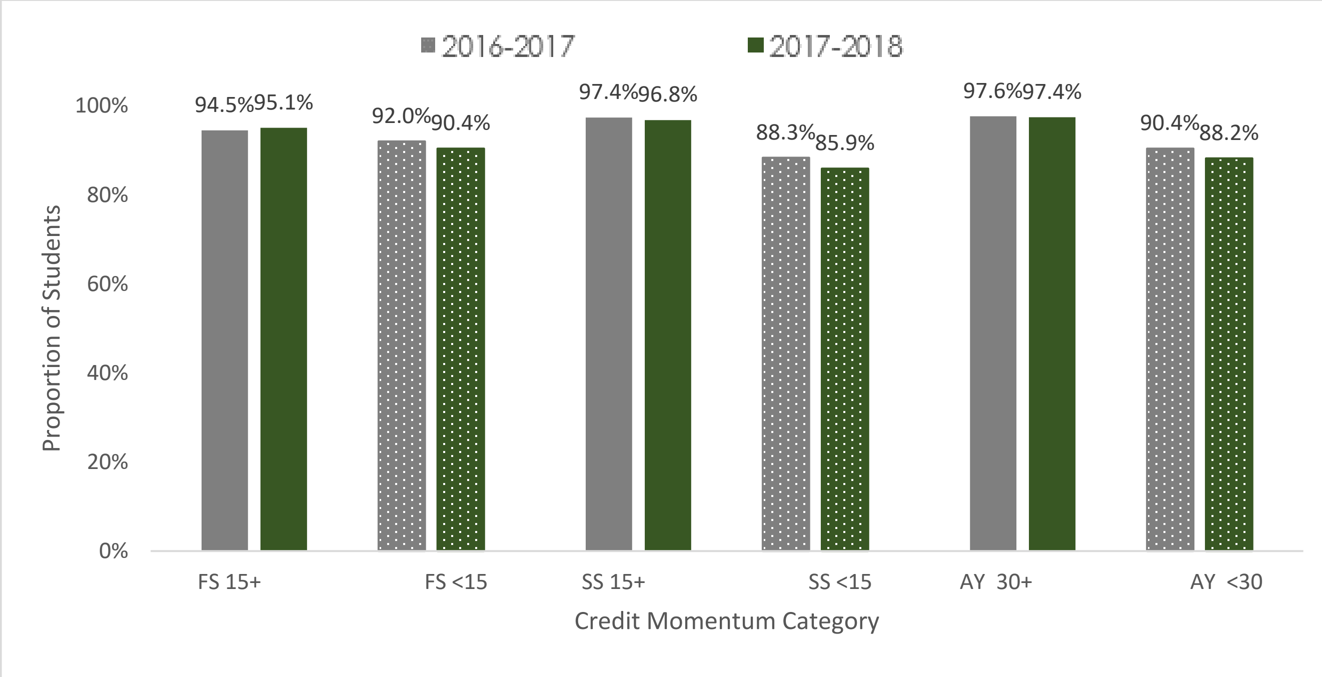

Figure 3: Proporiton of students who are in good academic standing by credit momentum category

In short, higher rates of credit momentum corresponded to higher rates of good academic standing. When MSU substantially increased the percentage of its first-year students who attempted 15 or more credits in their first semesters or at least 30 credits in their first 12 months, a minimal decrease was observed in the percentage of students in the higher momentum category who were in good academic standing (97.6 percent to 97.4 percent). Average cumulative GPA also remained steady for students who attempted 15 or more credits in the spring or fall and 30 or more credits in an academic year. For students who attempted fewer than 15 credits in the spring or fewer than 30 credits in an academic year, average cumulative GPA dropped by 0.1, on average.

The implementation of this campaign also identified course-scheduling difficulties, which were further hindering student success.

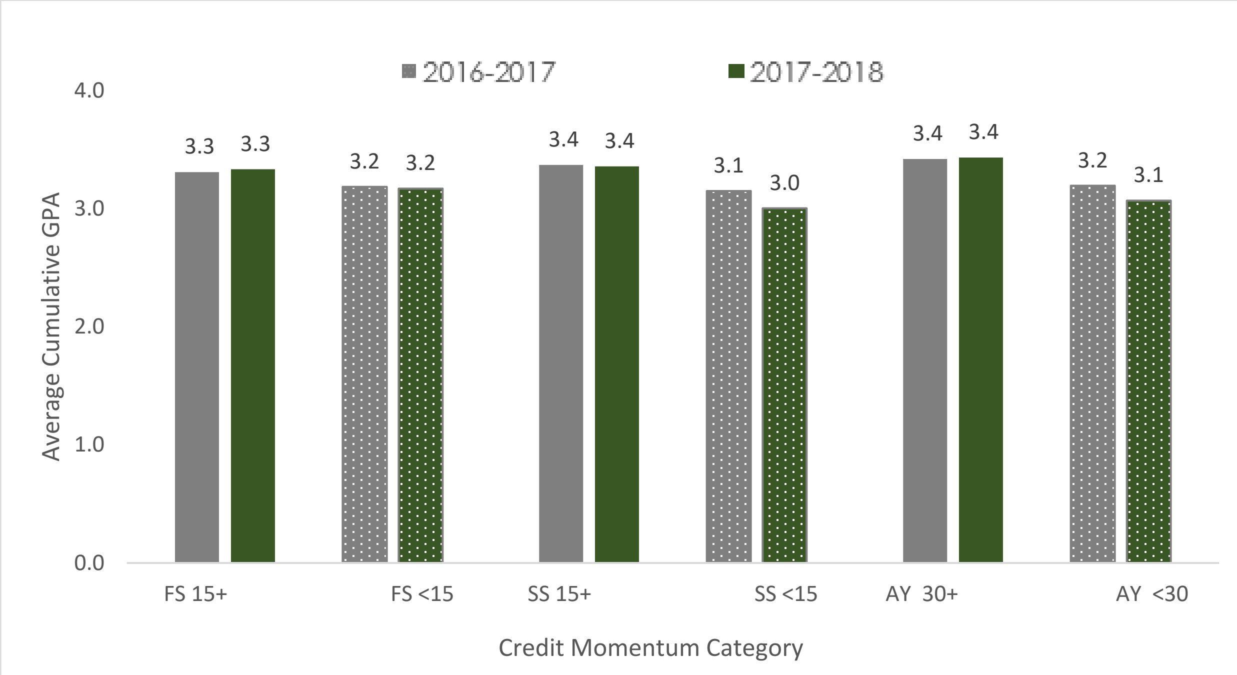

Figure 4: Average cumulative GPA on cohort and credit momentum category

Conclusions

This initiative demonstrated how an institution can use currently held data to implement student success policies and that collaborative work can be effective. Furthermore, use of commonly collected data can uncover challenges to student success and lead to improved understanding of students' course-taking behavior, including problems with the ways in which the institution scheduled courses. The found data enabled MSU to address barriers to student success without collecting more data.

References

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college.* Washington, DC: US Department of Education.

Attewell, P. & Monaghan, D. (2016). How many credits should an undergraduate take? *Research in Higher Education, 57,* 682-713.

Belfield, C., Jenkins, D. & Lahr, H. (2016). Momentum: The academic and economic value of 15-credit first-semester course load for college students in Tennessee. Community College Research Center Working Papers, 88, 1-35.

Bound, J., Lovenheim, M., & Turner, S. (2007). Understanding the decrease in college completion rates and the increased time to the baccalaureate degree. Report No 07-626. Ann Arbor, Michigan University of Michigan Population Studies Center.

Renata Opoczynski is the University Innovation Alliance Fellow at Michigan State University and can be reached at opoczyns@msu.edu.

Susan Richter is a Data Scientist in Institutional Research at Michigan State University and can be reached at srichter@msu.edu.

Bethan Cantwell is the Director of Institutional Research at Michigan State University and can be reached at cantwelb@msu.edu.

Mark Largent is the Associate Dean of Undergraduate Studies and Director of Learning Analytics at Michigan State University and can be reached at largent@msu.edu.

Comparing Assessment Data to Course Grades

By Lesley Page

Abstract

The department of Organizational Leadership at Lewis University collected data to compare student grades with assessment ratings for the same assignment. Results show that while strongly correlated, grades and assessment ratings are often based on unique criteria. However due to the close relationship between grades and assessment ratings, there is an opportunity to continue the discussion as to why and when research related to student grades may help those without access to direct student assessment data.

Introduction

For many of us working in higher education, our schools, colleges and universities continue to make progress with assessment. According to NILOA's recent trend report for 2018, "institutional needs for advancing assessment work have shifted since 2009 from engaging more faculty in assessing student learning to supporting faculty use of assessment results and wider stakeholder involvement" (Jankowski, Timmer, Kinzie & Kuh, 2018, p. 3). Nevertheless, there are still some faculty that do not buy into the process of assessment.

At faculty meetings, it is not uncommon to hear resistance to the University's need for assessment data, specifically:

- *Why should I have to do this?*
- *Why are you adding busy work to my plate when I'm not sure what the purpose of it is?*
- *Why can't we use grades as assessment data?*

Both long-time faculty who have never had to assess student learning before and newly hired faculty share in this skepticism, often feeling as though assessment may be more work than it is worth.
There are benefits to assessment that have been discussed extensively by leading authors in the field (Banta, 2002; Suskie, 2009). These include a focus on student learning, refined or enhanced curriculum, and continuous process improvements of both teaching performance and course content.

Yet many faculty who resist assessment efforts feel it is a personal evaluation of their own teaching performance that can lead to punitive consequences related to promotion and tenure, can infringe on faculty autonomy and academic freedom, and can diminish curricular control (Haviland, Turley & Shin, 2011).

With this in mind, faculty in the department of Organizational Leadership at Lewis University have been collecting data on student grades as compared to assessment ratings reported for the M.A. Organizational Leadership program hoping to be "myth busters" and debunk the belief that student grades can be a proxy for the assessment of student learning outcomes.

Background

In Lewis University's Master of Arts in Organizational Leadership (MAOL) program, faculty assess graduate student learning through the Capstone Course. As the last course in the program, students prepare a case study analysis (i.e., written paper) applying the concepts, principles and theories learned throughout the program. Consensus was reached by department faculty that the Capstone Course could effectively measure student learning across all six student learning outcomes (SLOs). Program level SLOs are listed below. Each SLO is theory based, relating to the conceptual ideas and principles covered in the program.

M.A. Organizational Leadership Student Learning Outcomes	
1.	Summarize foundational leadership theories which emphasize the dynamic relationship between leaders and followers as well as the influence of the leadership environment.
2.	Evaluate models which apply to the practice of leadership in areas such as organizational change, conflict management and team building.
3.	Explore the role of ethics in leadership.
4.	Incorporate critical thinking and decision-making skills to enhance leadership and organizational effectiveness.
5.	Evaluate the role of leadership as it relates to organizational culture, with consideration of issues related to individual and organizational differences.
6.	Use research to support decisions, especially related to organizational practices and improvement.

Table 1. M.A. Organizational Leadership student learning outcomes.

To assess student learning, a simple 4-point rubric is used that focuses on learning and application of each of the theory-based outcomes.

M.A. Organizational Leadership Assessment Rubric	
1=	Knowledge Not Demonstrated
2=	Basic Knowledge Demonstrated (no application)
3=	Basic Knowledge and Application Demonstrated
4=	Exceptional Knowledge and Application Demonstrated

Table 2. M.A. Organizational Leadership assessment rubric.

Comparing Assessment Ratings with Grades

This year the MAOL program gathered additional data to see how grades earned on the Capstone paper compared to assessment ratings. Since the same instructors teaching the Capstone Course were also conducting assessment, obtaining student grades from the Capstone paper was easy and efficient.

This process allowed the department to evaluate the art of grading vs. assessment. The art of grading involves reviewing the whole assignment to see how well the student answered the assignment prompts, used appropriate grammar and formatting, provided a clear and organized argument, demonstrated critical thinking and applied their learning. While assessment also looks at student performance and learning, usually the scope is comparatively narrower in nature (i.e., specific to SLOs). Therefore, the act of assessment often uses a rubric specific to mastery of the outcome, focused on varying levels of proficiency, and can be both skill and theory based (Ellis & Francl, 2015).

Key differences between assessment data and student grades are summarized in the following table.

Assessment of Student Learning Outcomes…	Grading of Student Assignments…
1. Focuses on specific learning outcomes and/or competencies gained from the course or program. 2. Demonstrates skill (e.g., written communication). 3. Combines theory and practice with a large focus on mastery of the learning outcome.	1. Focuses on assignment guidelines and requirements; ensuring all prompts of the assignment are answered. 2. Considers grammar, punctuation, and format (including APA or MLA). 3. Evaluates clarity and flow of the response. 4. Applies critical thinking skills. 5. Demonstrates application of theory and knowledge. 6. Integrates personal experience as it relates to class concepts and learning. 7. Includes participation, engagement and classroom behavior. 8. Includes attendance in class.

Table 3. Characteristics associated with assessment of student learning outcomes as compared to grading of student assignments.

These differences are also highlighted by comparing the 4-point assessment rubric rating scale (provided previously) with the grading rubric used for written assignments (see below), such as the Capstone paper.

Criteria for Evaluation				
Fulfills Assignment Requirements	**Formatting/Grammatical Accuracy**	**Clarity and Organization**	**Critical/Analytical Thinking**	**Application of Learning/Knowledge**
• Answers all prompts in the assignment • Meets assignment requirements in terms of page length	• Free of grammar and spelling errors • Properly formats sentences and paragraphs, including accurate punctuation • Properly uses APA format for citations and references	• Logical and clear organization of material • Writing is coherent and understandable • Reader is able to discern which prompt of the assignment is being answered • Coherently integrates cited material into the sentence	• Shows the ability to think critically (analytically) about the subject matter • Formulates and supports arguments • Shares the reasoning behind response or argument • Discusses or identifies different points of view	• Discusses and defines relevant theories, concepts, or principles • Analyzes how theories, concepts and principles from class can be applied (not only listing theories but discussing their application) • Synthesizes ideas/learning across perspectives or sources (which can include

			• Incorporates own point of view and/or personal perspectives, examples or ideas that support the argument(s) provided • Cites sources to support response/ argument	comparing and contrasting ideas or perspectives using personal examples to illustrate theories, concepts, or principles)

Table 4. Grading rubric used for written assignments in the MAOL program at Lewis University.

By reviewing the criteria included in the assessment rubric and grading rubric it can be predicted that assessment and grading do share some common aspects of student learning. Both focus on the learning demonstrated in coursework and application of knowledge. Given these commonalities, there should be a linear relationship between the two variables. And there is, as is evident in Figure 1.

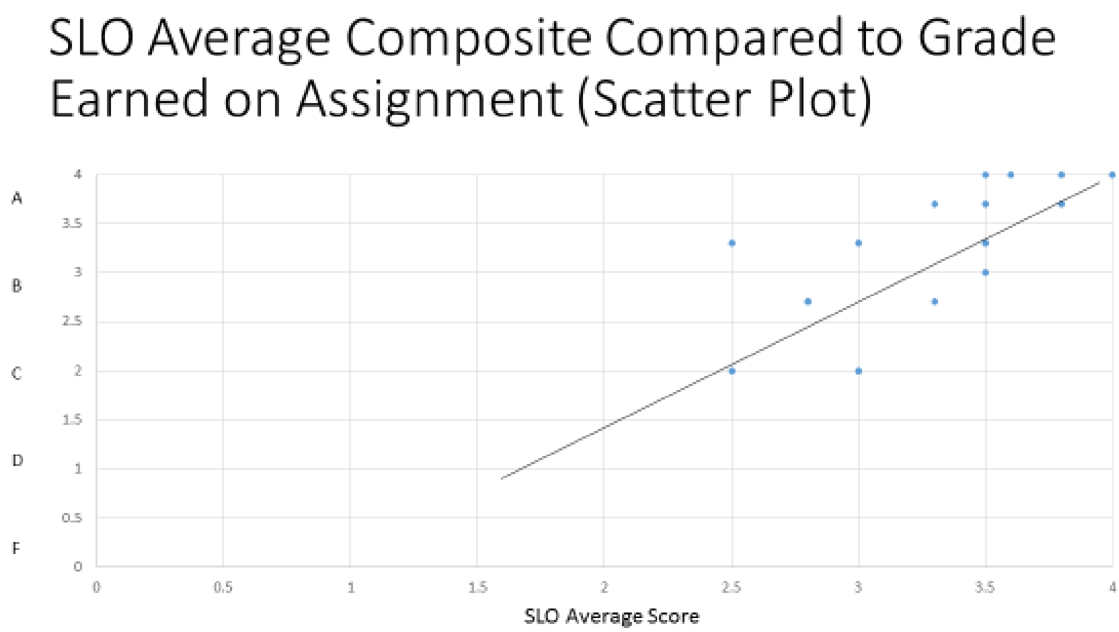

Figure 1. Student learning outcome composite compared to grade earned, r = 0.77, N = 22, with regression line plotted.

However, the differences in the rubric also point to clear distinctions as to how grading and assessment ratings are obtained. In other words, there are unique criteria used for grading than for assessment.

Assessment Rubric Average	Earned a Grade of…
3.5	A B+ B
3.3	A- B-
2.5	B+ C

Table 5. Comparison of SLO assessment rating composite with grades earned on the Capstone paper.

When we drill down into the specific grades earned, interesting data begins to emerge (see Table 5). We see that an assessment composite of 2.5, 3.3 or 3.5 (on a 4-point scale) may result in various grades ranging from A to C. This is because grading and assessment do not allocate scores or points for the same criteria. In this instance, grades are based not only on demonstrated mastery but also on class participation, grammar, flow/logic of the response and ability to think critically. These factors are not evaluated through assessment, which results in a variety of different grades being assigned to the same assessment SLO composite.

Due to small sample size and limited sample demographics (graduate students only), additional data would help to verify these conclusions across student populations, locations and learning modalities (online, blended, face to face, etc.).

Conclusions

When faced with deadlines, faculty ask if grades can serve as their assessment data of student learning. The data gathered by the department of Organizational Leadership at Lewis University helps to reinforce the argument that while similar in nature, assessment ratings and grades are not interchangeable. While we do see there is a linear relationship between grades and assessment ratings, the two are not the same. The difference is likely due to a difference in construct between the two measures.

Grades are appropriate for instructors to provide students immediate feedback on all parts of their assignment including fulfillment of assignment guidelines, grammar, logic/flow of written communication and critical thinking skills. Assessment, in comparison, does also help to identify critical thinking skills but focuses more specifically on student learning related to a competency or demonstration of theoretical/content knowledge. Assessment also engages faculty in the process of continuous learning and improvement. After all, as educators we share in the goal of fulfilling our promises to students in terms of what they can learn in our programs and offering coursework that is relevant to their learning. This is one of the reasons why assessment can be so powerful (Banta, Suskie & Walvoord, 2015).

While this study is limited in scope, an important implication is that in cases when assessment data may not be available or feasible, student grades can provide some basic analysis and help reach conclusions about student learning more broadly. This opens the door for further conversation on this topic.

References

Banta, T.W. (2002). *Building a scholarship of assessment*. San Francisco, CA: Jossey-Bass.

Banta, T.W., Suskie, L. & Walvoord, B.E. (2015). Three assessment tenors look back and to the future. *Assessment Update*, 27 (1).

Ellis, J.L. & Francl, T.J. (2015). Grading and assessments: Correlations of variables affecting teaching and course assessments. *Contemporary Issues in Education Research,* 8 (2).

Haviland, D., Turley, S., & Shin, S.H. (2011). Changes over time in faculty attitudes, confidence and understanding as related to program assessment. *Issues in Teacher Education*, 2011, 20 (1).

Jankowski, N.A., Timmer, J.D., Kinzie, J. & Kuh, G.D. (2018, January). *Assessment that matters: Trending toward practices that document authentic student learning*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Suskie, L. (2009). *Assessing student learning: A commonsense guide*. San Francisco, CA: Jossey – Bass.

Lesley Page is an Associate Professor of Organizational Leadership at Lewis University and can be reached at pagele@lewisu.edu.

Teachers' Dispositions in Action and edTPA

By Holly Thornton and Paige Neroda

Abstract

The topic of assessing educator dispositions continues to be an area of developing research. The Stanford Center for Assessment, Learning, and Equity (SCALE) developed the edTPA performance assessment for novice educators to demonstrate readiness. There are 784 Educator Preparation Programs in 40 states and the District of Columbia using edTPA as an assessment of culminating candidates, and thus a part of program assessment. This article examines an exploratory study of edTPA in relation to research on teacher dispositions. Dispositions in Action (Thornton, 2018) is a framework that examines how teacher dispositions affect learning and evidences what dispositions "look like" in practice. Findings indicate significant correlations between participants with higher responsive Dispositions in Action scores and higher edTPA scores. This implies that teacher preparation should examine intentionally "teaching" or cultivating responsive dispositions, just as they do other pedagogical skills thought to improve one's chances of success on the edTPA and as a future educator.

Introduction

The topic of assessing educator dispositions continues to be an area of developing research. Both theory regarding the nature and impact of educator dispositions on student learning and instruments to assess dispositions continue to evolve. Theory suggests there is a correlation between educators' dispositions and teaching proficiency. This paper seeks to discover if analyzing samples of graduate certificate interns (n=15, n=9) bears this out. With the sample size of this cohort group, only a strong match between the data and this theory is likely to be evidenced. A significant correlation would then lead to designing further study into the relationship between dispositions and effective teaching.

EdTPA has become a mandated part of our university's teacher preparation program, yielding data related to candidate proficiency in planning, instruction, and assessment. The Stanford Center for Assessment, Learning, and Equity (SCALE) developed the edTPA performance assessment for novice demonstrate readiness to teach through planning lessons to engage students in learning. Analysis of student learning grounded in best practices and related research happens within a reflective commentary portfolio process supported by video of candidate teaching (edTPA, 2018). There are currently 784 Educator Preparation Programs in 40 states and the District of Columbia that participate in edTPA, with 18 states requiring successful completion as part of their teacher licensure policy (AACTE, 2018). This performance data is then collected and archived for licensure requirements and program revision.

It may be beneficial to intentionally embed the edTPA's constructs of effective classroom planning, instruction, and assessment as meaningful elements within one's teaching and research. For candidates in a graduate certificate alternative licensure program, making connections between an online teaching environment and real schools through an edTPA lens can help candidates see the big picture and purpose behind the analytic and reflective thinking necessary to be successful on the edTPA. This also helps them to view teaching as decision-making and to ground the decisions that they make in solid research and best practices. From a research perspective, examining the edTPA in relation to research on teacher dispositions further builds on the potential benefits of using this data.

Data Source: edTPA

Thoughtfully embedding the edTPA learning targets as part of instructional design for a graduate certificate *curriculum and instruction* course can teach preservice educators how to think about and cultivate meaningful student learning. It helps preservice teachers, who come from backgrounds other than education, make sense of the complexities of teaching. The three sections of the edTPA are aligned with specific course activities and assessments.

The first section, planning, is easy to navigate. Candidates critique multiple approaches to lesson planning and engage in backwards design to construct engaging lessons. They analyze these lesson plans in terms of depth of learning using numerous taxonomies, such as Webb's Depth of Knowledge (2002) and the SOLO taxonomy (Biggs and Collis, 1982). The focus is on designing research-based lessons that enable students to demonstrate their proficiency in meeting lesson objectives and learning targets.

EdTPA's second section is the instructional section. This section requires candidates to analyze real teaching. Since the certificate students do not have the opportunity to engage in teaching in their own classrooms prior to their culminating internship, this has been a challenge. Clinical educators in our partner schools recorded and uploaded lessons to our online course site. The candidates then viewed the videos and used the commentary questions from edTPA to examine each teacher's decisions and provide contextual evidence and support of the candidates' analyses from the videos. Following this analysis, each clinical educator met online with the candidates to discuss their instructional and management decisions and to answer questions. This process led to collective deep thinking about teaching practices.

For part three of the edTPA, candidates need to engage in assessment. Candidates can design and critique assessments without having to set foot in an actual school; however, a main part of the edTPA assessment task is to give meaningful feedback and suggestions for future learning for real students. After watching and analyzing clinical educators' lessons, the candidates receive student work samples from said lessons to evaluate. The students then relate these evaluations to each lesson's goals and give specific, focused feedback to individual students based on student misconceptions, successes, and depth of understanding. By reviewing and evaluating actual student work samples from the clinical educators' lessons, teacher candidates learn how to give in-depth feedback to specific students, to analyze data across the class to reflect on successes, and to determine next steps needed for responsive instruction.

Concerns have been raised about the high stakes nature of the edTPA potentially leading to reductionist, test-driven approaches to teacher preparation (Greenblatt & O'Hara, 2015**)**. However, the aforementioned approaches allow edTPA to become an intentional and meaningful part of course design and instruction. These approaches also provide multiple course-embedded means of formative assessment and data gathering related to candidates' performance on teaching tasks. By the end of the course, teacher candidates' assessments comprise an entire "practice" edTPA commentary. This is evaluated using the edTPA rubrics and scoring system. By the end of the semester, multiple views on candidate success across the elements of planning, instruction, and assessment are available for unit reports and program evaluation. This data is useful in identifying trends in strengths and weaknesses related to each of the edTPA rubrics across the graduate certificate program.

Candidates' teacher dispositions are also evaluated with a collection of tools, resulting in multiple scores during the semester. The Dispositions in Action (DIA) evaluation protocol is used to enable preservice teachers to identify their dispositions. After reading about, analyzing, and discussing DIA, candidates write a self-evaluation using that model. The DIA word preference scale is also completed by the teacher candidates to provide another data point related to their dispositions. Additionally, candidates who are in their final semester of the program and teaching in a public school classroom are able to engage in a DIA assessment by their faculty supervisor using the observation tool.

Beyond its primary purposes, how can edTPA data be used? Does the edTPA data correlate with other teacher behaviors or characteristics, such as teacher dispositions? To examine this, the DIA data were compared to the edTPA data.

Data Source: Dispositions in Action

Research into DIA (Thornton, 2018) can give us insight into the impact of dispositions on teaching and learning in the classroom. It provides a framework that allows us to examine how teacher dispositions affect learning and what they "look like" in practice. Rather than focusing on prevalent approaches to teacher dispositions in the field, such as using professional behavior checklists, self-reflective journaling, hypothetical case analyses, or setting up a data system to document standards evaluation for accreditation, DIA examines how dispositions are evidenced in the classroom through teacher/student interactions (Thornton, 2006b). DIA reflects two emergent categories of dispositions: responsive and technical (Thornton, 2006a & Thornton 2006b). The disposition to be responsive is a thinking-based orientation that is responsive in many dimensions: responsive to the needs and actions of the learner, as well as their developmental characteristics, understanding, student questions, student work, and the learning context. Teachers with technical dispositions exemplify the role of teacher as technician, knowing how to successfully employ the skills of teaching, but not highly valuing or examining the "why" behind their instructional decisions. Instead, the focus is on efficiency and accountability. There is often little variation from situation to situation and student to student (Thornton, 2018).

Evidence of DIA can be aligned with major classroom functions, or domains, where they are typically exhibited. These domains of practice include instruction, assessment, and management, as described in the chart below.

Responsive Dispositions	Classroom Domain	Technical Dispositions
The disposition to be **Critical** in one's thinking. Evidenced in dialogue that is: probing, focused on quality, centered on criteria, concerned with deep understanding The disposition to be **Challenging** in one's thinking. Evidenced in dialogue that is: centered on high expectations, student competence and success for all students	Assessment	The disposition to be **Assuming** in one's thinking. Evidenced in dialogue that is: centered on completion of tasks, focused on correctness, concerned with grades The disposition to be **Accepting** in one's thinking. Evidenced in dialogue that is: indicative of low expectations, focused on effort and compliance
The disposition to be **Facilitative** in one's thinking. Evidenced in dialogue that is: guiding, inquiry oriented, concerned with application and connections to students' lives, and real-world	Instruction	The disposition to be **Directing** in one's thinking. Evidenced in dialogue that is: about directing actions of students, coverage of facts, telling information and giving answers

examples, in search of multiple answers and the exchange of ideas The disposition to be **Creative** in one's thinking. Evidenced in dialogue that is: about multiple ways of framing learning, examples, and paths to understanding diverse learners, responsive to students' questions, comments		The disposition to be **Repetitive** in one's thinking. Evidenced in dialogue that is: lacking in variety in explaining, exemplifying or representing learning, repetitive, the same way for all students
The disposition to be **Empowering** in one's thinking. Evidenced in dialogue that is: concerned with student input related to classroom instructional decisions, centered on fairness and equity The disposition to be in **Connected** one's thinking. Evidenced in dialogue that is: centered on developmental needs, exhibits "withitness" problem solving, conflict resolution and responsiveness to students as individuals	Management	The disposition to be **Controlling** in one's thinking. Evidenced in dialogue that is: concerned with managing student behaviors and actions including movement, talking, and other forms of interaction The disposition to be **Distanced** in one's thinking. Evidenced in dialogue that is: often limited, general in nature, generic, often remaining the same from class to class and situation to situation

The teacher candidates' DIA was evaluated using scaled scores from multiple assessments based in their *curriculum and instruction* course and on observation of their final internship teaching. The question this edTPA and DIA data could examine was, "Is there a relationship between these preservice teachers' Dispositions in Action and their edTPA scores?"

Previous research conceptually connected DIA with various teaching frameworks, such as the work of Charlotte Danielson (2007) and Lee Shulman (2004). In addition, ongoing qualitative observational data had indicated an emergent theme linking responsive dispositions to students who performed well on practice edTPA tasks. Our found data, a convenience sample from two existing graduate certificate cohort groups, enabled an exploratory study of this accessible population to occur and lead to the development of a study of the target population: initial certification students across multiple programs. These two independent sources of data could begin the examination of the correlation between dispositions and success on the edTPA, which claims to demonstrate teaching proficiency. DIA is one of the first models to connect dispositions with teaching practices and student learning outcomes. Examining potential connections between a standardized, required "proficiency" assessment (the edTPA), and assessment of teacher dispositions (DIA) is a next step in exploring whether a relationship exists between dispositions and teaching proficiency.

Findings
Each of the DIA assessments have items divided into and correlated with the areas of assessment, instruction, and management. The edTPA rubrics are divided into three sections reflecting three tasks: planning, instruction, and assessment. Total scores and sub-sections of these tools were analyzed using Pearson's correlation process. The first data set included fifteen participants who were in the semester before their final internship and examined scores using the "practice" edTPA. The second data set came from nine interns in the schools who completed the official edTPA.

The first candidate data set had a moderate positive correlation ($r=0.56$) between participants' scores on the dispositions word preference tool, which uses a five-point Likert scale, and the EdTPA planning scores. Participants whose dispositions were determined to be more highly responsive on the instructional word choice scale had higher scores on the planning rubrics (1-5) of the edTPA.

Examples of Responsive Choice/Technical Choice Items

- creative/reliable
- experimental /verified
- asking/telling
- explore/complete
- varied/consistent

The responsive choices (highly rating terms such as *creative*, *experimental*, *asking*, *explore*, and *varied*) appear to align conceptually with the planning tasks and rubrics in the edTPA. The planning task evidences how the candidates designed lessons and justified the choices they made. In edTPA planning rubric one, candidates are to exemplify a learning focus on inquiry, interpretations, or analyses and the use of supporting arguments and conclusions. Rubric two focuses on strategic thinking, the needs of specific individuals or groups, and identifying and responding to key student misconceptions. Responsive choices within items on the DIA word preference scale, such as *experimental*, *explore*, and *varied*, conceptually relate to these first two rubrics. EdTPA rubric three makes connections to research and/or theory, including student developmental characteristics. The preference of responsive word choices, including *creative*, *explore*, and *varied*, are well represented in the research base regarding responsive practices (Thornton, 2018). Rubric four considers the varied needs of students' language demands, including the type of discourse needed, which aligns with the respondents' preference for the term *asking*, which is indicative of dialogue between teachers and learners, rather than the didactic concept of *telling*. The last of the edTPA planning rubrics is about the use of multiple forms of evidence to monitor students' depth of understanding and allowing individuals or groups with specific needs to engage in appropriate ways to demonstrate their learning. This requires teachers to be both *varied* and *creative* in meeting all students' needs.

Stronger correlations were found in the second data set that compared the DIA to edTPA scores for nine graduate level interns during their final semester of full time field experience. The data indicated four significant correlations between the DIA observation scores and scores on the edTPA. Participants who received a higher overall responsive disposition observation score had higher overall edTPA scores ($r=0.69$, $N = 9$, $p < .05$). A second more focused correlation was found within the instruction-based sections in both the DIA observation and edTPA. The participants with higher scores in the DIA instruction domain showed a strong, positive correlation with edTPA scores on the instruction task rubrics($r=0.807$, $N = 9$, $p < .01$). Two other correlations were found within DIA observation scores and the edTPA planning task section and rubrics (1-5). A strong positive correlation was found between scores in the DIA instruction domain and scores on the edTPA planning task and rubrics ($r=0.795$, $N = 9$, $p = .01$). Further scores in the management domain were strongly correlated with the edTPA planning task ($r=0.795$, $N = 9$, $p = .01$).

Data Set Two Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
DIA Word Preference	9	.80	2.60	3.40	3.0000	.31623
DIA Word Preference: Instruction	9	2.00	1.80	3.80	2.8889	.61734
DIA Word Preference: Management	9	2.00	2.00	4.00	3.1778	.69602
DIA Observation: Assessment	9	2.00	1.00	3.00	2.1289	.82395
DIA Observation: Instruction	9	2.00	1.00	3.00	2.1100	.73192
DIA Observation: Management	9	2.00	1.00	3.00	1.9211	.82904
edTPA Planning	9	2.10	2.40	4.50	3.0333	.77460
edTPA Instruction	9	1.00	2.00	3.00	2.5333	.34641
edTPA Assessment	9	1.80	1.80	3.60	2.8000	.47958
DIA Observation Total Score	9	32.00	16.00	48.00	32.3333	12.51998
edTPA Total Score	9	19.00	31.00	50.00	41.2222	6.24055
Valid N (listwise)	9					

Discussion

Over the past twenty years of research conceptual connections between the Dispositions in Action model and multiple best practice frameworks have been found (Thornton, 2018). This early research on the potential relationship between candidates' Dispositions in Actions and their success on the edTPA indicates correlations worth further examination. Although this exploratory study includes only a small sample of participants, the level of significance found in the sample indicates interesting trends and the potential for further study.

Teacher education programs focus on developing skills related to successful completion of the edTPA. In fact, course embedded standardized assessments, often referred to as key or signature assessments, are often designed (or revised) around what edTPA "wants" candidates to do (Tuck & Gorlewski, 2016). However, it is possible that more than technical teaching skills affect teacher's success on the edTPA. For example, one's dispositions are the filter through which an educator views and lives in the classroom (Harris & *Sass*, 2011). Responsive Dispositions in Action correlate with instruction leading to deeper student understanding (Thornton, 2006a), as evidenced using the SOLO taxonomy (Biggs and Collis, 1982). Teacher candidates with responsive dispositions may be better able to evidence the complexities and grounding behind instructional design and classroom practices; this is key to success on the edTPA. A larger scale study of the relationship between Dispositions in Action and edTPA scores across programs is the next step.

The potential impact of a preservice teacher's dispositions on the design process and commentaries within the edTPA should be considered as another data point to inform teacher preparation. If dispositions correlate with edTPA scores, specifically more responsive dispositions correlate with higher scores on the edTPA, this implies that we should be just as concerned about how we "teach" or cultivate dispositions within a program as we are with other skills thought to improve one's chances of success on the edTPA. The definition and assessment of educator dispositions has long been problematic (Choi, Benson & Shudak, 2016). This practice-embedded definition of dispositions, DIA, allows assessing dispositions to move beyond self-reporting. behavior based checklists, or subjective evaluation of one's character or personality. This beginning study of the relationship between Dispositions in Action and edTPA provides an opportunity to examine how we might explicitly teach dispositions in our programs. It further emphasizes the need to intentionally cultivate responsive dispositions within our teacher candidates and interns, just as we would cultivate technical skill in the use of best practices. Teacher educators can no longer assume that dispositions are simply inherent in individuals or implicitly part of teacher preparation, nor can they continue to be treated as an add-on to courses for the sake of data reporting. The potential relationship between responsive dispositions and teaching proficiency should give us pause as we consider the importance of intentionally teaching candidates more than technical skills.

References

AACTE (2018). Retrieved from: http://edtpa.aacte.org/state-policy. June 15, 2018

edTPA (2018). Retrieved from: https://www.edtpa.com/PageView.aspx?f=GEN_AboutEdTPA.html June 15, 2018

Biggs, J. B. and Collis, K. (1982) Evaluating the Quality of Learning: the SOLO taxonomy. New York, Academic Press.

Choi, H., Benson, N. & Shudak ,N. (2016). Assessment of Teacher Candidate Dispositions Evidence of Reliability and Validity. Teacher Education Quarterly. Summer p71-89

Greenblatt, D. & O'Hara, K. (2015) Buyer Beware: Lessons Learned from EdTPA Implementation in New York State. Teacher Education Quarterly, 42 (2) p57-67.

Thornton, H. (2006a). Dispositions in action: Do dispositions make a difference in practice? Teacher Education Quarterly. 33(2), 53-68.

Thornton, H. (2006b). The disposition to teach against the grain: Responsive teachers as the key to the future of middle school. Gateways to Teacher Education. 18(1), 18-31.

Thornton, H.(2018). The it factor: What makes a teacher great? Sense/Brill , Boston, Tuck, E & Gorlewski, J. (2016). Racist ordering, settler colonialism, and edTPA :A participatory policy analysis. Educational Policy. 30 (1).

Webb, N. (March 28, 2002) “Depth-of-Knowledge Levels for four content areas” unpublished paper

Holly Thornton is Professor of Middle Grades Education at Appalachian State University and can be reached at thorntonhj@appstate.edu.

Paige Neroda, is a graduate student at University of North Carolina Charlotte and can be reached at pjthornt@uncc.edu.

Save the date for **AALHE 2019** at the InterContinental Hotel in Saint Paul, Minnesota

Association for the Assessment of Learning in Higher Education

AALHE

PROMOTING ASSESSMENT FOR LEARNING

INTERSECTION

Editorial Staff

Intersection is the quarterly publication of the Association for Assessment of Learning in Higher Education | 60 Terra Cotta Ave., Suite B #307, Crystal Lake, IL 60014 | 859-388-0855 | info@aalhe.org