Embedded Tagging and Radar Map Shape Analysis for Assessing Student Outcomes

Dr. Gabriel R. Burks, University of Illinois, Urbana-Champaign

Gabriel Burks is a postdoctoral research associate at University of Illinois Urbana-Champaign. He is from Louisiana where he also completed his Baccalaureate Degree in Physics with a minor in Chemistry from Grambling State University. Dr. Burks received his doctoral degree from Drexel University in Materials Science & Engineering and was also a National Science Foundation GK-12 STEM Fellow for 3 years, in the School District of Philadelphia. Primary research interests are: 1) synthesis and characterization of functional polymeric materials, 2) polymer crystallization phenomena, and 3) pedagogical assessment of conventional and nontraditional STEM spaces.

Dr. Jennifer R. Amos, University of Illinois, Urbana-Champaign

Dr. Amos joined the Bioengineering Department at the University of Illinois in 2009 and is currently a Teaching Associate Professor in Bioengineering and an Adjunct Associate Professor in Educational Psychology. She received her B.S. in Chemical Engineering at Texas Tech and Ph.D. in Chemical Engineering from University of South Carolina. She completed a Fulbright Program at Ecole Centrale de Lille in France to benchmark and help create a new hybrid masters program combining medicine and engineering and also has led multiple curricular initiative in Bioengineering and the College of Engineering on several NSF funded projects.

Colin Widmer Castleberry, University of Illinois, Urbana-Champaign
Abstract

Computer-based assessment has been shown to offer many benefits on the outcomes of student performance [1]. The computational strengths of computer-based platforms allow for more in-depth collection and analysis of data from students of today than the many years previous. This access to data and performance outcomes allow us to learn more about the individual student instantly and to use this feedback to tailor adapted teaching and learning experiences on the fly. Here a locally developed computer-based testing platform is used to administer, assess and collect stealth identification tagged data from quizzes, homework, and exams of undergraduate students in sophomore-level engineering course. The information gathered from this platform is represented using an accessible radar plot format then analyzed using a novel method, based on the shape of the radar plot, to develop a greater understanding of the individual strengths and deficiencies of students. Finally, to establish appropriate context for this data it is correlated to a common student success metric and analyzed for potential trends.

Introduction

Computer-based assessment has been shown to offer tremendous benefits for student performance outcomes [1]. As classrooms become more and more diverse, it becomes more imperative that classroom environments are sufficiently equipped to handle the demands of educating scholars who possess equally diverse modes of thinking and learning [2-4]. The call for diversity in STEM and increased numbers of people obtaining degrees has resulted in classrooms with higher populations of women and ethnic and racial minorities [5, 6]. In many cases, the onus is placed on a singular instructor to carry such a burden of identifying and remaining consistently sensitive to the increased diversity in their classroom. Beyond the nontrivial task of providing equally diverse pedagogy, developing a grasp of the complex social and behavioral environment that exist in a typical classroom ecosystem is even more challenging [7-10].

Buried underneath the theme of diversity exists the need to identity specialized talent to service national and societal workforce interests. As society and job offering become more and more specific, a skilled workforce with equally specific skillsets must be trained and identified as potential candidates to fill said positions. Several assessments exists in military and industry spaces to identify talents possessing specialized skillsets; however, these identification-based assessments have room to be incorporated more into tradition classroom environments [11-13].

Computer-based learning [14-18] and assessment [1, 19-23] platforms are uniquely suited to ease the pressures placed on instructors to sufficiently teach and track such diverse groups of learners and to facilitate the identification of a talented 21st century workforce. The integration of
computers into everyday learning environments and the increased processing power of computer networks allows educators to collect more data than ever on today’s students. More data, though it may be extremely useful, could also easily be considered a hindrance to education progress being that in many cases information overload and without appropriate context can be categorized as practically useless [24, 25]. Today there is a significant need to analyze complex data, while also considering the simplicity of the representation of these data systems.

In this work, computer-based assessment and testing are used in conjunction with a novel data analysis technique to develop a potentially more in-depth learning profile for undergraduate engineering students. The presented insights may shed light on easily addressable student deficiencies and provide a point of departure for instructors to better tailor their course content and to provide a more efficient learning experience for diverse student populations. Secondary to the acquisition and analysis of computer-based assessment data and in an attempt to provide appropriate context for the data presented through this novel method, this suite of nontraditional assessment data is analyzed for trends against proven education standard data.

**Methods**

In this study, the University of Illinois Urbana-Champaign (UIUC)-developed PrairieLearn (PLN) platform is used to administer, assess, and collect student performance data in a sophomore level engineering course. Prior to implementation of course, a set of relevant and identifying tags are produced to appropriately label assessment questions. The setup of the PLN platform and development of appropriate tag list can be quite onerous and was done with the assistance of a former course instructor and graduate student.

Besides PLN for assessment tasks, data was collected and processed using a standard data analysis software. No special equipment or software was necessary to process and produce the data and figures presented in this work.

**Results and discussion**

**Integrated tagging assessment**

Through the PLN platform specific identification tags are embedded into quizzes, homework assignments, and exams for students. These tags may represent core content concepts covered within the specific question or fundamental skills, identified by the instructor, that are deemed necessary to solve the question. To begin analyzing the pool of students, it is important to understand the impact that these tags have on the expected outcome for each question. As shown in Figure 1, student question success may be related to the number of tags represented in a given question.
This tag density is related to the core complexity of a given question. For example, a question with one tag could be thought of as being very straightforward where if the student understands that core concept or skill then they should have minimal difficulty in solving that question successfully; to the contrary, if a question has multiple concept or skill tags, then a student would have to possess the ability to both understand the individual concepts or skills as well the ability to see the problem in a global fashion to systemically apply all of the knowledge. Figure 1 provides a strong indication that as the tag density in a given question increases, then the student success rate for that given question subsequently decreases. This is true for skill and concept tags operating independently and for the tags operating in combination.

This information alone, for instructors insistent upon using such assessment could provide a means for determining the limit for the incorporation of core concepts into a given question to maximize student learning retention. Secondary to this idea of potentially overloading a question with too rich learning content is that the instructor may also get a better understanding for how the number of times a particular core concept or skill is utilized within a given assessment period and how the repetition of tags impacts student performance. Preliminary results in that space seem to indicate that the repetition of tags over the period of a course should result in increased proficiency and performance for students.

**Introduction to roughness and radar shape-based assessment**

As a means to begin to reconcile with such a large amount of data and a wide variety of data variables, representation of specific skill or concept data in the form of radar plots was put forth. An example of a typical radar plot, incorporating student skill tags are shown in Figure 2.
For a given radar plot, each axial ray represents the proficiency of a singular student in the specific skill area. This proficiency is represented on a scale of 0 to 1 and is to show the normalized student score on all questions that incorporate the given skill tag as a percent fraction of 1 (100% proficiency). For Figure 2, this particular student has a proficiency of roughly 0.63 or 63% for all questions that incorporate the skill tag “algebraic equation,” and 1.0 or 100% proficiency for questions that include the skill tag “statistics.” The nine skill tags that will be referred to in this study are: 1) algebraic equation, 2) differential equation, 3) integral equation, 4) trigonometry, 5) unit conversion, 6) statistics, 7) system integration, 8) system of equations, and 9) dimensional analysis.

This form of data representation provides a clear and global view of a given students’ performance on the singular assessment level and even across the level of an entire course, if adjusted to do so. It is possible to better understand how the proficiency in any given skill evolves throughout a course as content is introduced and formally assessed within a classroom.

In this work, the use of radar plots and this particular representation of data and student performance was of particular interest, with the primary question being, how can we use this plot to understand more about a given student beyond the context of a singular assessment or singular course term? Drawing inspiration from traditional engineering methods to determine the sphericity of droplets or spherical particles, a similar analysis is initiated to determine the roughness of student radar plot profiles [26]. To conduct this analysis, the roughness, $R_a$, and root-mean-square roughness, $R_{RMS}$, are calculated based on the values for each axial ray on given student plot profile. The equations for $R_a$ and $R_{RMS}$ can be seen in Equation 1 and Equation 2, below.
Equation 1. Roughness, $R_a$, where $L =$ evaluation length and $Z(x) =$ the profile height function.

$$R_a = \frac{1}{L} \int_0^L |Z(x)| \, dx$$

Equation 2. Root-Mean-Square Roughness, $R_{RMS}$, where $L =$ evaluation length, and $Z(x) =$ the profile height function.

$$R_{RMS} = \left[ \frac{1}{L} \int_0^L Z(x)^2 \, dx \right]^{1/2}$$

When each radar plot profile is viewed as a potentially perfect sphere, 0 roughness, the roughness value is a quantitative and physical means of representing how close to a perfectly smooth sphere a given spherical object actually is. Both $R_a$ and $R_{RMS}$ are representations of “surface” roughness, and both calculate some variance of the average peaks and valleys associated with a given surface, which may or may not be spherical. Sphericity is not a requirement for surface roughness measurement. $R_a$ and $R_{RMS}$ differ in that $R_{RMS}$, by virtue of its equation calculation method is more “sensitive” to the effects of individual axial ray peaks and valleys, which would result in a higher value for roughness when compared to $R_a$.

In this work, both $R_a$ and $R_{RMS}$ are calculated and utilized for analysis. This type of analysis of student-based data has not been done before, therefore it is important to showcase both $R_a$ and $R_{RMS}$ values in an attempt to inch closer to a foundational context for this given type of analysis as a research community.

In addition to the $R_a$ and $R_{RMS}$, the average circular radius, $r_{AVGCIR}$, is also calculated. The average circular radius is simply the average of all axial rays divided by two. The intent of the average circular radius is to provide some representation for the total level of proficiency reached by a given student profile, for example an $r_{AVGCIR}$ of 0.5 would represent 100% proficiency across all skill tag sectors and an average circular radius of 0.25 would represent an overall average skill proficiency of 50%. In context, this would mean that a given student, when all axial rays are combined and regardless of specific axial ray values, has proficiency of 50% of all the skill concepts presented via assessment.

Without explicitly showing the individual calculations for roughness, beyond the given equations, Figure 3 is an example of a set of students who represent the four extremes of this analysis, with respect to their classroom peers. In each example from Figure 3, A-D, each radar plot includes the global student radar profile, a classification of smooth or rough, a level of proficiency, a classification of “Well-Rounded” or “Specialized,” and a value representing their respective total score in the class on PrairieLearn based assessments.
Figure 3. Four extreme scenarios for radar plot profiles analyzed for roughness and average circular radius or proficiency.

For Figure 3A, this particular student profile represents the largest roughness value among the classroom cohort. Larger values of roughness means that the given student has large peaks and low valleys in their skill-based proficiency. As it can be seen in 3A, this student has very high levels of proficiency on questions that include the skill tags for statistics and trigonometry, whereas they have very low levels of proficiency in assessment questions that include the skill tags dimensional analysis and system integration. To begin to place roughness into some appropriate context, the descriptor of “Specialist” or “Specialized” are used. These descriptors would indicate that the given student has very strong talents in a few specific skills at the expense of proficiency in other skills. Figure 3B represents the student with the largest average circular radius value, which means that that student has the highest level of average proficiency across all skill tag questions compared to the class. Though this student does not represent the lowest roughness value, when compared to others this student can be considered fairly “well-rounded”. A well-rounded profile may be likened to an individual that is good at all skills, but a master of none. This may not be the case when the average circular radius very high in addition to the roughness value being very low, indicating a high level of individual skill mastery along
with very few semblances of peaks and valleys, respectively. Figure 3C represents the student profile with the lowest roughness value or “smoothest” radar profile. It is interesting to note that the most well-rounded student profile does not necessarily equate to the highest proficiency or highest grade on assessments. This particular student is above-average proficient in every skill tag category, but has a class average of 76.2 vs 81.5 for the student in 3B. The final extreme scenario, can be found in Figure 3D. The student in Figure 3D represents the lowest \( r_{AVGCIR} \) value among the data set. This profile is comprised of many low valleys and a singular large peak. Unsurprisingly, this student also has a lower total grade for the course at 40.8. This specific profile indicates a student who is struggling to grasp most fundamental skills necessary to be successful in the question assigned within the assessment questions assigned by the instructor.

Beyond the individual student radar profiles, it is important to establish a view of the spread of students within the entire group set. The breakdown of students based on their respective roughness and average circular radius values are shown in Figure 4, below.

![Figure 4](image)

**Figure 4.** Class breakdown of root-mean-square roughness and average circular radius. Axis set at class average for each.
The average $R_{\text{RMS}}$ and $r_{\text{AVGCIR}}$ for the class are 0.16 and 0.33, respectively. Using the class average values as the x- and y-axis points and without any other points of reference, Figure 4 may be viewed as a potential representation of how this particular class shapes out with respect to the extreme scenarios presented in Figure 3. The upper left quadrant of Figure 4 represents a well-rounded (low roughness) student with high proficiency (high $r_{\text{AVGCIR}}$); the upper right quadrant represents a specialized (high roughness) student with high proficiency (high $r_{\text{AVGCIR}}$); the lower left quadrant represents a well-rounded (low roughness) student with lower proficiency (low $r_{\text{AVGCIR}}$); and the lower right quadrant represents a specialized (high roughness) student with low proficiency (low $r_{\text{AVGCIR}}$). A mock example of what the profiles should look like for the most extreme scenarios in each quadrant can be seen in Figure 5.

![Figure 5. Global and specialist proficiency mock scenario quadrant for students with high levels of proficiency and developing proficiency.](image)

Echoing the real student examples shown in Figure 3, Figure 5 is a mock diagram for what is to be expected in each quadrant for respective student radar profiles when analyzing for roughness and global assessment-based skills proficiency. The descriptors, Global/Holistic, represent analogs for “well-rounded” and are intended to indicate that students who are within this classification have demonstrated that they have learned or developed proficiency in the assessed skill tags in a global or balanced fashion. To the contrary, the descriptors, Analytic/Specialists, represent analogs for “rough/specialized” in that they are intended to indicate that the students
who are classified as such are really strong in some things and comparatively weak in others. It is important to stress that it should not be considered a negative or positive to be classified in either the Global/Holistic or Analytic/Specialist sectors. Just as any research lab would have instrument technicians who specialize in singular suites of instruments or lab managers/principal investigators who have a more global view of the research enterprise, there is a strong necessity for individuals who are classified by both sectors.

**Relationships between grade point average and roughness assessment**

The following section, highlights some potential relationships between a student’s individual radar plot profile roughness, average circular radius, and their cumulative grade point average (gpa). The assumption being that students with a more global or a well-rounded grasp of the fundamental skills would be more likely to perform better outside of the engineering classroom. This would mean that a student with a lower roughness value would have higher cumulative gpa than a student with a higher roughness value. Figure 11A-11C show trends in roughness, proficiency, and students’ cumulative grade point average.

![Figure 6. Relationship between A) $R_a$ and student gpa, B) $R_{RMS}$ and student gpa, and C) $r_{AVGCIR}$ and gpa.](image)

The first global comparison is to be made between roughness, $R_a$, and root-mean-square roughness, $R_{RMS}$, as it relates to grade point average. Because there is minimal reference point for what roughness means in an education or student context, the comparison between roughness and know values is necessary to begin establishing that context. At a fundamental level, it is
necessary to make appropriate determination for which roughness calculation method provides the most relevant information within this education space. As it may be seen in Figure 11A and 11B, both roughness calculations are inversely proportional to grade point average. Both figures show that students with lower roughness values tend to have higher cumulative grade point averages. It is still early in this study and the slope for both, though negative, still represent a modest trend in this system. This first look at an association between well-rounded or holistically learned individuals to cumulative grade point average seems to indicate a clear trend.

As it relates to the student proficiency and cumulative grade point average relationship, shown in Figure 11C, there appears to be a much stronger trend whereby student with higher $r_{AVGCIR}$ values tend to be higher gpa performing students. This trend seems slightly more understandable in that it also means that students who do well in this class, tend to do well other classes. High performing students tend to be high performing all the time.

Summary

The incorporation of tag-based markers into computer-based assessment was used to analyze how specific student skill-based proficiency may be related to other aspects of student performance. It has been shown that it is possible to simplify complexed student-based assessment data using nontraditional data representation and analysis through the PrairieLearn assessment platform. This advanced analysis has been used to identify student skill deficiencies and strengths, which may allow for instructors to adjust content offerings mid-term to fit the skill-based needs of the class or individuals. This data may be further parsed to provide breakdowns for specific population subgroups and to better understand potential behavioral trends within a classroom ecosystem. Finally it was shown that it may be possible to develop context for this unique analysis method and its resulting variables (roughness and $r_{AVGCIR}$) by investigating relationships between the new experimental values and values of known relevance such as grade point average.

Such computer-based assessments will prove to provide a more fair and equitable assessment of diverse student populations and their respective demographic subgroups. A better understanding of the individual components that comprise a student, especially when presented in a simple and easily digestible manner, is surely to improve the instruction and learning experiences for all students.

Acknowledgements

This work is supported by NSF IUSE/PFE:RED - #1623141 and covered under University of Illinois IRB #14864.

References


