Tracking undergraduate student achievement in a first-year physiology course using a cluster analysis approach

S. J. Brown, S. White, and N. Power

School of Interprofessional Health Studies, Auckland University of Technology, Auckland, New Zealand

Submitted 22 May 2015; accepted in final form 3 August 2015

Brown SJ, White S, Power N. Tracking undergraduate student achievement in a first-year physiology course using a cluster analysis approach. *Adv Physiol Educ* 39: 278–282, 2015; doi:10.1152/advan.00076.2015.—A cluster analysis data classification technique was used on assessment scores from 157 undergraduate nursing students who passed 2 successive compulsory courses in human anatomy and physiology. Student scores in five summative assessment tasks, taken in each of the courses, were used as inputs for a cluster analysis procedure. We aimed to group students into high-achieving (HA) and low-achieving (LA) clusters and to determine the ability of each summative assessment task to discriminate between HA and LA students. The two clusters identified in each semester were described as HA (n = 42) and LA (n = 115) in semester 1 (HA1 and LA1, respectively) and HA (n = 91) and LA (n = 42) in semester 2 (HA2 and LA2, respectively). In both semesters, HA and LA means for all inputs were different (all P < 0.001). Nineteen students moved from the HA1 group into the LA2 group, whereas 68 students moved from the LA1 group into the HA2 group. The overall order of importance of inputs that determined group membership was different in semester 1 compared with semester 2; in addition, the within-cluster order of importance in LA groups was different compared with HA groups. This method of analysis may 1) identify students who need extra instruction, 2) identify which assessment is more effective in discriminating between HA and LA students, and 3) provide quantitative evidence to track student achievement.

assessments; cluster analysis; nursing; undergraduate

**ASSESSMENT TASKS** in an undergraduate course are often used as the basis for progression into higher levels of study and possibly indicate whether or not a learning objective has been achieved by a student. Therefore, appropriate assessment tasks may guide student learning as it influences both the approach to learning (1, 9) and confirms the attainment of a learning outcome (10). Formative assessment may provide a student with feedback to support the achievement of a learning outcome (15) and may provide feedback in a nonthreatening environment (3), as this type of assessment usually has no course credit assigned to it (13, 14). In contrast, summative assessment is primarily used to grade students (for example, at the conclusion of a study period) and often provides no feedback to students on their performance. Scores achieved in summative assessments are often emphasized by both educators and students, and performance in these assessments may be the decisive factor of a student’s progression (8).

In the present report, we describe the novel use of a data classification technique to identify groups of students based on their summative assessment scores. This clustering analysis, which is a form of data mining, identifies clusters embedded in data, where a cluster is a group of data objects that are similar to one another (4, 7, 16). The similarity of objects in a cluster may not be obvious and may be based on a collection of measures rather than a single criterion. Although the number of variables used to define a group are not predetermined, a sufficient number is needed to ensure adequate discrimination between items placed in a group. In the present report, we attempt to group students using a two-step cluster analysis, using their academic achievement scores in all summative assessment tasks.

In an introductory anatomy and physiology course, enrollment prerequisites, suitable content, and appropriate assessment tasks help to ensure relevant learning outcomes are achieved by most students (11, 12). However, evolution of the course consistent with evolving new pedagogies (18) also necessitates retrospective course evaluation by educators on the course. These reflective practices may review the methods of assessment used on a course to ensure the chosen tasks are fit for purpose; it is possible that a retrospective cluster analysis performed on student summative assessment scores will provide information suitable for such a review. Therefore, in this study, we aimed to 1) identify groups of high-achieving (HA) and low-achieving (LA) students, and 2) track student achievement in assessment between the two semesters.

**METHODS**

**Setting.** This study was carried out at a large, publicly funded university in regional Victoria, Australia. The university has a long tradition of preregistration undergraduate nurse education, preparing students for state registration. The two human anatomy and physiology courses were compulsory for all students, and each was taught over a 13-wk semester, in a student’s first undergraduate year. The first year comprised two semesters, with full-time students expected to take four courses in each semester. Each week, students attended a 2-h lecture and a 1-h lecture (on separate days) and were required to attend a weekly 1-h tutorial. Students were encouraged to spend ~10 h/wk on independent study and were expected to purchase an appropriate human anatomy and physiology textbook. Although attendance was not compulsory, a register of attendance was kept and students were expected to provide appropriate documentation to explain absences.

**Data collection and analysis.** Data were accessed from the university’s data management system with the approval of the course coordinator. Throughout the analysis, deidentified, aggregated data were used, thus presenting no student privacy issues. This study did not require a full submission to the university’s Ethics Committee, although appropriate advice was sought from both the university research advisor and the university privacy officer; a condition was that no individuals could be identified by the researchers, nor could a student identify their own data from the analysis.

The two-step cluster analysis is an exploratory strategy designed to reveal natural groupings (or clusters) within the data set that otherwise would not be apparent. The two-step method has the advantage that no
a priori allocation of the number of clusters is required and that the importance of each input variable for the construction of a specific cluster is identified. The method standardizes all input variables but does not allow a missing value for any input variable.

The range of marks available for each input was 0–10 for all progress tests (PTs). There were three PTs in the semester 1: the first PT (PT1) in week 4, the second PT in week 8 (PT2), and the third PT (PT3) in week 12. This pattern was repeated in semester 2, with the first PT (PT1_2) in week 4, the second PT (PT2_2) in week 8, and the third PT (PT3_2) in week 12. The range of marks available for both midsemester multiple-choice tests was 0–50, and midterms were carried out in week 6 of semester 1 (Midterm) and again in week 6 of semester 2 (Midterm_2). The range of marks available for both final exams was 0–80, and these was carried out either 1 or 2 wk after the final teaching period in week 13, depending on timetabling restrictions. The semester 1 exam (Exam) and the semester 2 exam (Exam_2) had the same format, whereby 80% of the available marks were allocated to multiple-choice questions and 20% was allocated to handwritten short answers. Alphabetic grades were derived from the final numeric score, where a C (49.5–59.5), B (60–69.5), and A (70–100) denoted a passing grade and a D (<49.5) denoted a failing grade.

All data were analyzed using SPSS (IBM SPSS Statistics 22). Each numeric score for the five assessments was used as an input variable in the cluster analysis. All scores were considered as continuous variables. All inputs were standardized. For each input, the data interval (minimum, maximum) was converted to an interval (0, 1) using the following formula: new value = (original input value − minimum) × (1/(maximum − minimum)). This conversion, such that all input data had a maximum possible value of 1 and a minimum possible value of 0, allowed comparisons between each input and between semesters. In addition, all inputs were equally weighted by the two-step clustering technique, the number of clusters was determined automatically, and the distance between variables for cluster allocation was determined using the log-likelihood method. Deidentified data were available for analyses such that no individual student could be identified from examination of the values. Before these data were released, the assessment scores for each student were entered into a spreadsheet and ordered by a unique student identification number; this ensured that when the data were entered into the cluster analysis, there was minimal order effect.

Data for students in semester 1 were analyzed separately from data obtained in semester 2 but using an identical method. Within each semester, clusters were compared with either an independent-samples t-test (for inputs of Exam and Midterm) or a Mann-Whitney U-test (for inputs of PT1, PT2, and PT3). A nonparametric test (Mann-Whitney U-test) was used to compare the PTs between the clusters as these scores ranged from 0 to 10 and may have not satisfied the distributional assumptions of a parametric test (e.g., independent-samples t-test).

This report only includes data on students who completed all assessment tasks in both courses and achieved an overall pass mark in both courses (a final summative score of >49.5%). It does not include those students who did not complete all of the assessment tasks in each course, and it does not include any students who completed all assessments in the semester 1 course but did not progress into the semester 2 course; this would include a student who withdrew from the program or a student who attempted all assessments but did not achieve a high enough final summative score to pass the semester 1 course. In this institution, students who do not pass a compulsory course that is a prerequisite for a following compulsory course are counseled by the program leader and may be offered an alternative academic pathway.

RESULTS

Data from 157 students were included in the analysis. This was the total number of students passing both courses, and no students who did not pass the semester 1 course progressed into the semester 2 course. A typical intake into a semester 1 course for this program would be 200 students. Two clusters were identified in both semesters. In the semester 1 course, the LA group (LA1) contained 115 students and the HA group (HA1) contained 42 students. In the semester 2 course, the LA group (LA2) contained 66 students and the HA group (HA2) contained 91 students. The analysis allowed the tracking of students in both semesters, such that 19 students moved from the HA1 group into the LA2 group in semester 2, 68 students moved from the LA1 group in semester 1 into the HA2 group in semester 2, and 70 students remained in the same group (either LA or HA) in both semesters.

Overall importance of inputs was different in semester 1 compared with semester 2. In semester 1, the final exam was determined as having the highest importance followed by the midterm test. In semester 2, the midterm test was determined as having the highest importance followed by the final exam. Within each semester, the importance of inputs within each of the clusters was also different; these are shown in Table 1. For each semester, the clusters were significantly different (all P < 0.01) when each input was compared between LA and HA; these data are shown in Table 2. A comparison between the clusters and the total group for semester 1 and semester 2 is shown in Fig. 1.

Alphabetic grades were used to indicate the achievement of students on both courses. After completion of the course in semester 1, the LA1 cluster contained 38 “B” grade students and 77 “C” grade students, whereas the HA1 cluster contained 40 “A” grade students and 2 “B” grade students. After completion of the course in semester 2, the LA2 cluster contained 4 “A” grade students, 23 “B” grade students, and 39 “C” grade students, whereas the HA2 cluster contained 82 “A” grade students and 9 “B” grade students.

DISCUSSION

In the present study, we uniquely report the use of a two-step cluster analysis procedure to retrospectively examine undergraduate nursing students’ academic achievement in two introductory human anatomy and physiology courses. The novel

Table 1. Order of importance for inputs in the two-step cluster analysis

<table>
<thead>
<tr>
<th>Order of importance</th>
<th>LA1</th>
<th>HA1</th>
<th>LA2</th>
<th>HA2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Exam</td>
<td>Midterm</td>
<td>Exam_2</td>
<td>Midterm_2</td>
</tr>
<tr>
<td>2</td>
<td>Midterm</td>
<td>Exam</td>
<td>PT2_2</td>
<td>PT1_2</td>
</tr>
<tr>
<td>3</td>
<td>PT2</td>
<td>PT3</td>
<td>PT3_2</td>
<td>PT2_2</td>
</tr>
<tr>
<td>4</td>
<td>PT3</td>
<td>PT1</td>
<td>PT1_2</td>
<td>PT3_2</td>
</tr>
<tr>
<td>5</td>
<td>PT1</td>
<td>PT2</td>
<td>PT2_2</td>
<td>PT1_2</td>
</tr>
</tbody>
</table>

LA1 and LA2, low-achieving clusters in semesters 1 and 2, respectively; HA1 and HA2, high-achieving clusters in semesters 1 and 2, respectively; PT1, PT2, and PT3, first, second, and third progress tests in semester 1, respectively; PT1_2, PT2_2, and PT3_2, first, second, and third progress tests in semester 2, respectively.
findings reported here are that on these courses, a clear distinction can be made between HA and LA students and that different assessment tasks were more capable of discriminating between HA and LA students. The method is also capable of tracking students who may move between groups in each semester.

In Fig. 1, the median, upper, and lower quartiles for each input variable are shown as clear boxes, and, therefore, Fig. 1 allows a comparison of these values with the median and data spread (colored points and whiskers) of each cluster. In semester 1, the HA cluster median values were consistently above the upper quartile for the total group, whereas in semester 2, the HA cluster median was consistently between the total median and upper quartile. This indicates that the academic achievement of the total group more closely matched that of the HA group in semester 2 compared with semester 1. We speculate that this may reflect a progression in the difficulty of the content delivered in semester 2 and perhaps a predisposition to teach more material aimed at challenging the more able students. Alternatively, the content delivered in the second course may have required an understanding of content delivered in the first course, thereby advantaging those higher-achieving students who were more capable at coping with the further development of physiology concepts. However, it could be argued that a progression in content difficulty would result in lower scores for many students; this was not supported in the present study, as the mean scores for each assessment (shown in Table 2) were comparable between semesters.

A different picture emerges when the performance of the LA group to the total group was compared. In semester 1, the LA cluster median for all inputs was consistently between the median and the lower quartile for the total group, whereas in semester 2, the LA cluster median was either at, or lower than, the lower quartile for the total group. This may suggest that the LA cluster was slipping further behind (in academic achievement) relative to the total group, perhaps further supporting the notion that material in semester 2 was more challenging. These observations highlight the value of our cluster analysis approach as a technique to provide empirical evidence to support the reflective practices of a course coordinator. This retrospective analysis may be used to highlight the need for appropriate scaffolding of content; this is to ensure that concepts in physiology are clearly presented with an increasing level of difficulty, suitable for the students enrolled on this type of course (introductory anatomy and physiology).

Reflection on student performance in assessment can be an important part of course development and help determine future iterations of the course. As part of this reflection, an educator may wish to identify which of the assessment tasks were best able to discriminate between higher- and lower-achieving students and which of the assessment tasks were found to be the most challenging by the lower-achieving students. In both courses in this research, the weightings of each component were 30% for the PTs, 30% for the midterm test, and 40% for the final exam. A student was expected and encouraged to treat each assessment point with
clearly discriminate between students into HA and LA clusters. Although this could explain why this input was rated as having high importance by our cluster analysis, we suggest that this is not the case. In the courses presented in this study, the style of questioning was similar in all assessment tasks (predominantly either multiple-choice or single word/statement answers), with an allocation of only 20% of the final exam to handwritten, short answer-style questions, and even these could be adequately answered by some bullet points or a labeled diagram. Therefore, we suggest that the identification of the exam as a dominant input by which to separate students into HA and LA clusters is a valid finding.

Classifying students according to their performance in assessment tasks is traditional within academia, providing a means to rank students with regard to their peers, national standards, or the level to which learning outcomes have been achieved. For educators, it may be useful to identify which assessment task is most likely to distinguish between a high achiever and a low achiever. Also, if a number of assessments are used to calculate a final summative score for a student, an educator may choose to allocate a weighting for each assessment based on its ability to discriminate between HA and LA students. In the present report, we demonstrate that the cluster analysis technique can assign an importance to an input based on its ability to discriminate group membership, i.e., an input that more clearly separates high achievers from low achievers is rated as “more important” than an input which is less likely to discriminate between high and low achievers.

The cluster analysis did not take into account a “pass” or “fail” grade for any input, meaning that the overall attainment of a pass for the course (a combined summative score of >49.5%) was not a criteria for inclusion into a cluster. Although a high pass rate may satisfy some requirements for future progression within the university, it may have limited use in course evaluation, planning, and the progressive evolution of the course. Thus, we suggest that the cluster analysis, as described in the present study, is a more useful mechanism by which student performance on a course can be evaluated.

Some aspects of these two compulsory courses in physiology were perceived as difficult by some students, a finding consistent with other similar courses (17, 19). Others (5, 6) have stated that knowledge of physiology was perceived by health professionals as important, essential for questioning medical decisions, and ensuring patient safety, but was limited in its undergraduate delivery.

Nursing undergraduates have previously reported that they were generally disappointed with the lack of integration of physiology theory with practice (2). In the present study, the LA cluster may indicate that some of the undergraduates who pass the courses are still in need of extra instruction in anatomy and physiology and that some undergraduate nursing students may continue to struggle with anatomy and physiology content. While this lower achievement may be multifactorial, we speculate that the lack of specific nursing application of the physiology content, which may apply physiological concepts to nursing practice, may be a contributing factor. For example, undergraduate nursing students may learn physiology best when it is related to workplace practices (2), although recently qualified nurses have reported that anatomy and physiology in their preregistration (undergraduate) curriculum was rarely linked to practice. Although speculative, it may be that those
students in the LA cluster would benefit more from the presentation of physiology content when the nursing application of such content is more overt.

Limitations. In the present study, the cluster analysis is not intended to replace the use of grades, but it can offer a course instructor an additional lens to examine student academic performance and progression. Alphabetic grades are commonly used to describe academic achievement, and in the current courses students who passed were awarded either A, B, or C grades. The numeric range that determined the alphabetic grade could be adjusted by the course coordinator following recommendations by an oversight committee; however, in our analysis, the grade cutoffs were consistent with university policy.

This cluster analysis required all inputs from both semesters; thus, it only identified students who may need extra instruction after both courses are completed. However, in identifying these students at an early stage in an undergraduate pathway (with possibly 3 more years of study to complete), it could be possible to implement strategies in later courses that may provide this instruction. Students who pass but are consistently in the LA cluster may be restricted in their choice of electives or offered an alternative sequence of courses. For example, a second-year course that develops an understanding of the physiological principles underpinning pharmacology could be offered to HA students in semester 1 and to LA students in semester 2 or a physiology revision course could be implemented for LA students in semester 1.

In this cluster analysis, only assessment scores were input variables, such that a student’s score in each assessment task was used as the input variable for the cluster analysis, and this provided five inputs to quantify group allocation. Our future research will attempt to combine other psychometric measures, for example, scores for student attitude and engagement, with assessment scores in a similar cluster analysis; this may identify further groupings (e.g., a LA group with low scores for both attitude and engagement).

The present study only clustered and tracked students through two undergraduate courses in anatomy and physiology, where the consistency of assessment tasks in each course made this more straightforward. Future application of the technique may use assessment scores from other courses throughout an entire undergraduate curriculum, thus providing a tool to track students’ progress through their chosen degree pathway.

DISCLOSURE

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: S.J.B., S.W., and N.P. conception and design of research; S.J.B. performed experiments; S.J.B. analyzed data; S.J.B. interpreted results of experiments; S.J.B. prepared figures; S.J.B., S.W., and N.P. drafted manuscript; S.J.B., S.W., and N.P. edited and revised manuscript; S.J.B. approved final version of manuscript.

REFERENCES