// Research Article //

Quantitative Measure of Student Retention of Information in Human Anatomy and Physiology: A Case Study

Jewel A. Daniel, PhD
jdaniel@ndm.edu

From the Biology Department, School of Arts, Sciences & Business, Notre Dame of Maryland University, Baltimore, Maryland.

QUANTIFYING INFORMATION RETENTION

Retention of information is essential for transfer of knowledge from one course to another. Human anatomy and physiology (A&P), offered as a 2-semester course at Notre Dame of Maryland University, is a foundational prerequisite for many health-related programs.

ABSTRACT

Retention of information is essential for transfer of knowledge from one course to another. Human anatomy and physiology (A&P), offered as a 2-semester course at Notre Dame of Maryland University, is a foundational prerequisite for many health-related programs.

For this study the researcher attempted to quantify the knowledge retention decline in the transition from human A&P I to human A&P II. Two cohorts of female traditional college students were administered a cumulative final exam immediately on completion of human A&P I.

One cohort (CS1) was given the same test 48 days later. A second cohort (CS2) was given the same test 48 days and 144 days later. There was a significant decline in retention of information in CS1, however, CS2 exhibited no significant decline at either 48 days or 144 days. Interestingly, there was no significant difference between both cohorts on the initial test, an indication that both cohorts were equivalently prepared. Further study is required to understand the disparity in retention decline between the 2 cohorts.

INTRO — DUCTION

Retention of information and skill is crucial for learning, whether the focus is kindergarten to 12th grade, or within higher education. Transfer of knowledge is critical for foundational courses whose information and skills are requisites for higher-level courses.

Retention of information and skill is crucial for learning, whether the focus is kindergarten to 12th grade, or within higher education. Transfer of knowledge is critical for foundational courses whose information and skills are requisites for higher-level courses. Many instructors anecdotally report a loss of knowledge in students transitioning from a lower-level course to a higher-level course.

A study of approximately 600 undergraduates studying biological sciences at 5 universities in the United Kingdom showed a significant decline in performance when given an A-level biology exam, a pre-university exam required for admission to biological sciences (Jones et al., 2015).

One explanation is that students learn as much as needed for the test and then forget the information after completing the test, a concept informally known as "cram and dump". Several school systems have made remedial efforts to limit this practice by placing emphasis on continuous or progressive testing, active learning, and higher-order thinking according to Bloom's taxonomy (Cuevas, 2016; Custers, 2010; Healy et al., 2017; Yielder et al., 2013). For example, Yielder et al. (2013) proposed progressive testing in Australian and New Zealand medical schools as a way to curb "cram and dump" and increase learning retention. A carefully crafted study by Healy and colleagues (2017) found that interrupting learning with carefully placed quizzes increased retention.

Efforts to quantify the learning loss have been made since the advent of institutional learning (reviewed in Semb & Ellis, 1994). Arthur Jr. et al. (1998) refer to learning loss as "skill decay" and define it as the loss of trained or acquired skills or knowledge over a period of time of non-use.

Via a meta-analysis of 189 data points from 53 articles published prior to 1998, they found that the amount of learning retention loss is proportional to the interval of non-use and is dependent on the type of information and skill (Arthur Jr. et al., 1998). In an effort to counteract the popular belief that most information gained in the classroom is forgotten immediately, Semb and Ellis (1994) conducted an analysis of 21 studies using a recognition test format and quantified learning loss or loss factor (LF) as 15% over a retention interval of 10 to 40 weeks post-instruction.

The formula RS = OS – (OS x LF), where RS is retention score, OS is original score, and LF is loss factor, can be used to predict the approximate performance on subsequent tests on the same subject matter (RS). Wisher et al. (2001) used this formula to distinguish between the RS of students using distance learning compared to in-class learning and found no significant difference.

With the advent of the COVID-19 pandemic and the subsequent global school shutdown in 2020, there has been a resurgence of studies quantifying learning loss (Donnelly & Patrinos, 2022; Hevia et al., 2022). It is important to note that in this case, learning loss is not equivalent to retention of information loss or "loss factor". Learning loss, as defined here, is the decline in collective student knowledge and skill and compares skill levels of current cohorts' testing to skill levels of previous cohorts at the same level of learning. Retention, as defined by Arthur Jr. et al. (1998) and in this current study, refers to the knowledge and skills retained by the same individual or cohort of students over time.

In higher education, retention of information and skills is critical in foundational courses because higher-level courses assume a level of competence based on these courses. For the biological sciences across higher education, key concepts in introductory courses such as Chemistry of Life, Evolutionary Theories, Cells and Cell Theory, and the many processes of the cell are foundational for most upper-level biology courses. Human anatomy and physiology (A&P) is a prerequisite for multiple health science programs and professions ranging from certificate courses to terminal degrees.

This includes but is not limited to respiratory technology, emergency medical technician, radiological science, nursing, physician's assistant, and occupational therapy programs. Medical schools offer anatomy in the first year. Thus, a solid foundation in human A&P is critical for the success of students pursuing health science degrees.

At many 2-year and 4-year colleges human A&P is offered either as a single semester course that samples the breadth of human organ structure and their related functions or a 2-semester course that further develops the structure and related functions.

Both the Community College of Baltimore County and Notre Dame of Maryland University (NDMU) offer a 2-semester (15 weeks/semester) human A&P course that is a prerequisite for entry into nursing programs. In both schools, enrollment in human A&P is contingent on successful completion of an introductory biology course. At NDMU, the first semester of human A&P (human A&P I) covers the foundational concepts of anatomical terminology, histology, the structural systems, and the regulatory systems and culminates in a cumulative final exam. Successful completion of human A&P I with a grade of C or higher is required for enrollment in human A&P II. Human A&P II covers the transport, exchange, and reproductive systems.

Two case studies were conducted at NDMU with human A&P pre-nursing students to assess learning retention decline in the transition between human A&P I and human A&P II. NDMU is a small private liberal arts university in Baltimore, Maryland with an ethnically and economically diverse student body. Human A&P is offered in the School of Arts, Sciences and Business which, at the time of this the study, was a women's undergraduate college. Students were arranged in cohorts with sections separated by pre-nursing and non-nursing. Over 90% of the students taking human A&P I in the fall semester enrolled in human A&P II in the subsequent spring semester.

The 2 cohorts in the study had the same instructor for both human A&P I and human A& P II. Instructions were given face-to-face with a combination of lecture and lab. The study measured only the lecture portion of the course. At the conclusion of human A&P I, students were given a cumulative final exam. Seven weeks and 20 weeks after instruction the same exam was administered, and the learning loss was calculated.

METHODS

Two case studies were conducted at NDMU with female students enrolled in human A&P. The course fulfills a prerequisite for nursing and only pre-nursing students are enrolled. The prerequisite for this class is successful completion of Fundamentals of Biology with a grade of C or higher. The course, taught via a systemic approach, is given over 2 15-week semesters with a cumulative final exam at the end of each semester. Emphasis is placed on structure, function, and pathology of tissues, organs, and organ systems.

Each cohort of students was kept consistent with greater than 90% of students transitioning from human A&P I to human A&P II. Enrollment in human A&P II requires successful completion of human A&P I with a grade of C or higher. The same instructor and the same textbook were assigned to each cohort adding another level of consistency.

In the first case study (CS1), 11 students, aged 19 to 22 years (modal age 19 years), were given a cumulative final exam at the end of human A&P I and the average performance of the class was recorded. Students were administered an identical exam 48 days later at the beginning of human A&P II after a 7-week break from instruction. In the second case study (CS2), 22 female students, aged of 19 to 26 years (modal age 19 years), were given the cumulative final exam at the end of human A&P I. The same exam was administered 48 days later at the beginning of human A&P II after a 7-week break from instruction, and then again 144 days later after completion of human A&P II. It is important to note that CS1 and CS2 were given the same exam questions which consisted of a combination of multiple choice, true or false, matching, short answer, and essay-type questions.

For all exams, questions that were graded subjectively, that is essay-type and short-answer questions, were removed from the exams and the average was tallied. Students who did not attempt all of the exams were also omitted from the average. Differences in average were analyzed using ANOVA and Tukey's HSD tests.

RESULTS

Case Study 1: The cohort in CS1 included 11 students who took both test 1 (T1, given day 0) and test 2 (T2, given day 48). The average score on T1 was 79.6% and ranged from 69.3% to 98% with a standard deviation (SD) of 7.94. The median score for T1 was 79.3 with a variance of 57.3. The average significantly decreased ($p < 0.005$) on T2 given 48 days post-instruction to 53.1 (SD 5.29) (Fig. 1a).

The scores ranged from 46.2% to 65.3% with a median score of 52.9 and a variance of 25.5. Individual scores for T1 and T2 are shown in Figure 1b. The difference between the T1 and T2 was determined to be statistically significant via ANOVA and Tukey's HSD ($p < 0.005$) (Table 1).

Table 1. Pairwise Comparison of Test Scores by Tukey's HSD.				
Pairwise Comparison	**Mean**	**Tukey HSD**	**Q**	**Significance**
CS1 T_1:T_2	M_{T1} = 79.57 M_{T2} = 53.05	26.53	Q = 7.91 (p = 0.00000)	*** p < 0.005
CS1 T_1:T_2	M_{T1} = 83.97 M_{T2} = 78.24	5.74	Q = 1.71 (p = 0.74622)	* ns
CS1 T_1:T_3	M_{T1} = 83.97 M_{T3} = 79.43	4.54	Q = 1.35 (p = 0.87339)	* ns
CS1 T_2:T_3	M_{T2} = 78.24 M_{T3} = 79.57	1.20	Q = 0.36 (p = 0.99909)	* ns
CS2:CS1 T_1	M_{CS2} = 83.97 M_{CS1} = 79.57	4.40	Q = 1.31 (p = 0.88553)	* ns
T_2 CS2:CS1	M_{CS2} = 78.24 MC_{S1} = 53.05	25.19	Q = 7.51 (p = 0.00001)	*** p < 0.005

T1 = Test 1 (day 0); T2 = Test 2 (day 48); T3 = Test 3 (day 144). * = no significant difference; *** = significantly different (p < 0.005).

Relative loss, described by Semb and Ellis (1994) as the amount of information remembered over a period of time, is calculated as the retention score by the equation RS = OS – (OS x LF). After evaluating over 21 studies, they determined an average LF of 15% over 10 to 40 weeks. Using this equation for CS1, the predicted average score on T2 is about 67.66%. However, the average score on T2 was actually 53.1%, less than predicted. Via simple subtraction calculation (T1 – T2), the calculated LF was 26.5%.

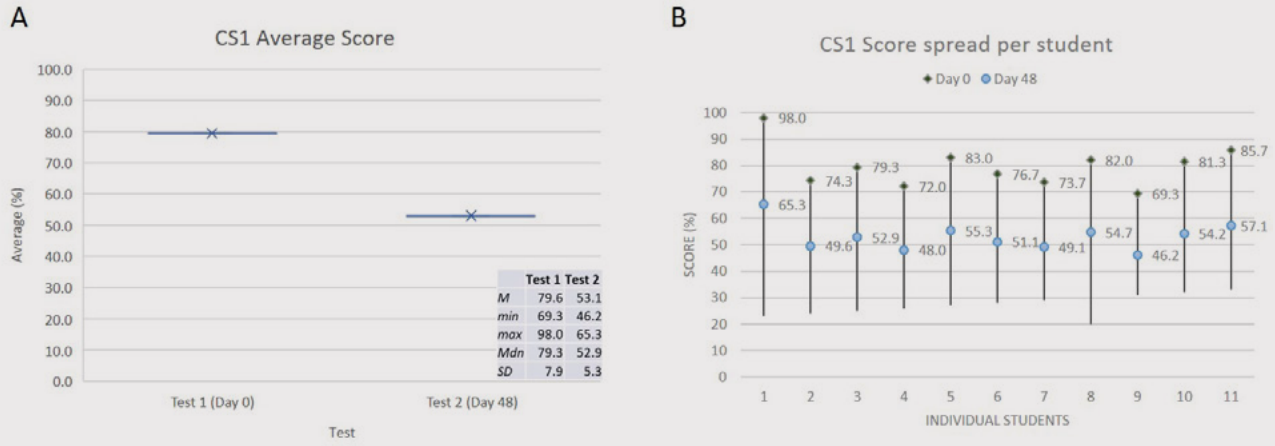


Figure 1. Comparison of tests for CS1 immediately after instruction (T1) and 48 days post-instruction (T2). Panel A, left, shows class averages for T1 and T2 (p < 0.005). Data table inset shows the mean (M), minimum score (min), maximum score (max), median (Mdn), and standard deviation (SD). Panel B, right depicts the score spread showing individual student performance on T1 (gray) and T2 (blue). Data labels show each student's scores on the tests.

CASE STUDY 2

22 students met the criteria for inclusion in the cohort for CS2. For this cohort, the retention score was measured both 48 days and 144 days post-instruction.

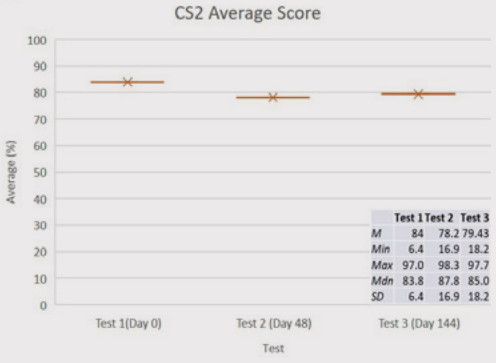
The students were administered the same test immediately after the completion of instruction (T1, given day 0), 48 days post-instruction (T2), and 144 days post-instruction (T3). Note that T3 was administered at the conclusion of human A&P II. Students only received instruction on topics covered in human A&P II.

Students in this cohort scored an average of 84.0% on T1 (SD 6.4), 78.2% on T2 (SD 16.9), and 79.4% on T3 (SD 18.2) (Fig. 2a). ANOVA and Tukey's HSD suggest there was no significant difference between T1 and T2 ($p = 0.74622$), between T1 and T3 ($p = 0.87339$), or between T2 and T3 ($p = 0.99909$) (Table 1).

Using the average LF of 15% as defined by Semb & Ellis (1994), one would predict a score of about 71.4% on T2 and T3, with T3 being lower than T2. Instead, the averages were 78.2% and 79.4%, giving a LF of 5.8% and 4.6% respectively, neither of which were significant. In fact, the LF between T2 and T3 was -1.4%, suggesting a gain of retention (not significant). Differences between individual student performance showed widespread variation (Fig. 2b).

Both case studies were given the same test and had the same instructor allowing for a direct comparison of both cohorts. Table 1 shows there was no significant difference between T1 between the cohorts, however, the average performance on T2 was significantly different between the 2 cohorts as was the average score for T1 of cohort 1 and T2 of cohort 2 ($p < 0.005$)

A



B

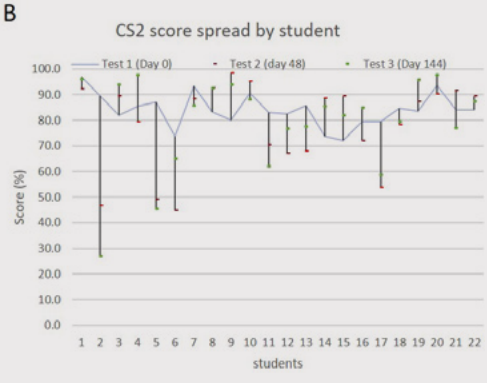
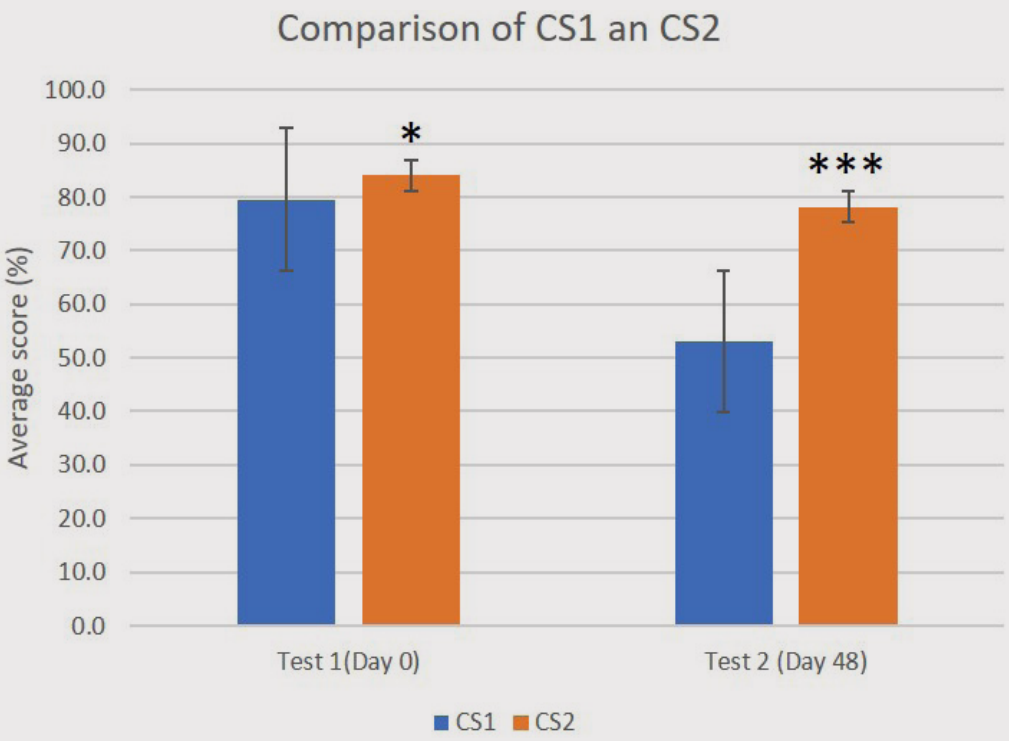


Figure 2. Comparison of test scores for CS2 immediately after instruction (T1) and 48 days post-instruction (T2) and 144 days post-instruction (T3). Panel A, left, shows calculated class averages on T1, T2, and T3.

Inset data table shows mean (M), minimum score (min) maximum score (max), median (Mdn), and standard deviation (SD). No significant difference exists between each data point. Panel B, right, depicts individual scores for each student on the 3 different tests. Blue line indicates scores on T1, brown dash indicates T2, and green dash indicates T3.

FIGURE 3.



	Mean	Tukey HSD	Q	Significance
Pairwise Comparison		$HSD_{.05}$ = 13.2317 $HSD_{.05}$ = 15.9598	$Q_{.05}$ = 3.9437 $Q_{.05}$ = 4.7568	
Test 1 CS2:CS1	M_{CS2} = 83.97 M_{CS1} = 79.57	4.4	Q = 1.31 (p = .88553)	* ns
Test 2 CS2:CS1	M_{CS2} = 78.24 M_{CS1} = 53.05	25.19	Q = 7.51 (p = .00001)	*** p < 0.005

Figure 3. Comparison of CS1 to CS2. Panel A, above, shows average scores of CS1 and CS2 on T1 and T2. Panel B, below, presents Tukey's HSD pairwise comparisons of CS2 to CS1 for T1 and T2. * = not significant (ns); *** = p < 0.005.

DISCUSSION & CONCLUSION

Retention decline or retention loss in students has been quantified by multiple studies (Arthur Jr. et al., 1998; Jones et al., 2015; Semb & Ellis, 1994; Wisher et al., 2001). There has been some inconsistency in the extent to which retention declines and the efficacy of corrective methods to minimize the decline.

The metadata analysis conducted by Arthur Jr. et al. (1998) and the retention loss test between distance and traditional learning performed by Wisher et al. (2001) involved recognition tests, requiring a low level of Bloom's taxonomy. Wisher and colleagues (2001) reported a retention loss of 14% to 16%, consistent with Semb and Ellis (1994), and showed no significant difference between the distance learning and traditional groups.

DISCUSSION & CONCLUSION

Many college courses, especially in the biological sciences, use a combination of recall, comprehension, application, analysis, evaluation, and synthesis, which incorporates higher levels of Bloom's taxonomy.

Consequently, a standard measure of retention decline is improbable to apply across college courses. However, the transfer of information is essential when transitioning from one course to another and that is dependent on student retention of knowledge.

The 2 case studies in this paper quantifying learning loss showed inconsistent results. Both of the cohorts in CS1 and CS2 were given identical tests. The performance on the first test, a cumulative final exam given at the end of human A&P I, was not significantly different between the cohorts. That indicates the 2 cohorts were equally matched in terms of understanding the material. Both cohorts met the same requirements for entrance into the course, that is a C or higher on the pre-requisite Fundamentals of Biology course. For both CS1 and CS2, the students were taking human A&P at the college level for the first time. Both were taught by the same instructor via a similar pedagogical approach.

What distinguishes the 2 cohorts from each other is their performance on the second test (T2) administered 7 weeks after T1. CS1 demonstrated a significant decline in retention after 7 weeks without instruction with an average of 26% lower score on T2 compared to T1. 100% of the students scored lower on T2 than on T1. However, CS2 exhibited no significant decline either in 7 weeks or 20 weeks after the initial test. Moreover, while the majority of individual students scored higher on the initial test than the second test, 8 of the 22 students (36%) scored higher on T2 than T1 and 12 (54%) scored higher on either T2 or T3 than T1.

There are several factors that were different about the cohorts that may contribute to the variation. The most obvious difference is the size of the cohorts. CS1 consisted of 11 students that took both human A&P I and II, while CS2 consisted of 22 students. Possibly a larger cohort size in CS1 would more reflect the results in CS2.

During their studies, both cohorts received interruptions in face-to-face instruction due to the COVID-19 pandemic. This interruption would have affected them at different stages of their education. While one cannot quantify or distinguish the effects of the disruption on either cohort, it's worth noting that studies have shown disparities in learning loss across different socio-economic lines due to the pandemic (Donnelly & Patrinos, 2022; Hevia et al., 2022). For those students who may have been dispersed in disparate high schools during the pandemic, the learning loss may be different to students who were already in the same college at the time of the shutdown. However, this is an unlikely explanation as students in both cohorts performed similarly on T1.

Human A&P, as offered, has a lecture component and a lab component. Data were only generated from the lecture component in this study. A possible reason for the disparity in retention decline between the cohorts may be the lab component. In CS2, the lab component was more application based with clinical case studies in addition to the identification of anatomical structures and function. In CS1, the lab emphasized anatomical structure and function with fewer clinically applicable case studies.

While the exact impact of the lab instructions on retention loss is beyond the scope of this study, studies indicate prior knowledge and knowledge gained outside the classroom have an impact on reducing retention loss (Semb & Ellis, 1994). More controlled studies are required to examine the effect of different modes of instruction on learning and retention. However, there is a lot to be learned from these case studies.

REFERENCES

1. Arthur Jr., W., Bennett Jr., W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11*(1), 57–101. https://doi.org/10.1207/s15327043hup1101_3

2. Cuevas, J. A. (2016). Cognitive psychology's case for teaching higher order thinking. *Professional Educator, 15*(4), 4–7.

3. Custers, E. (2010). Long-term retention of basic science knowledge: A review study. *Advances in Health Science Education, 15*(1), 109–128. https://doi.org/10.1007/s10459-008-9101-y

4. Donnelly, R., & Patrinos, H. A. (2022). Learning loss during Covid-19: An early systematic review. *PROSPECTS, 51*(4), 601–609. https://doi.org/10.1007/s11125-021-09582-6

5. Healy, A. F., Jones, M., Lalchandani, L. A., & Tack, L. A. (2017). Timing of quizzes during learning: Effects on motivation and retention. *Journal of Experimental Psychology: Applied, 23*(2), 128–137. https://doi.org/10.1037/xap0000123

6. Hevia, F. J., Vergara-Lope, S., Velásquez-Durán, A., & Calderón, D. (2022). Estimation of the fundamental learning loss and learning poverty related to COVID-19 pandemic in Mexico. *International Journal of Educational Development, 88*, 102515. https://doi.org/10.1016/j.ijedudev.2021.102515

7. Jones, H., Black, B., Green, J., Langton, P., Rutherford, S., Scott, J., & Brown, S. (2015). Indications of knowledge retention in the transition to higher education. *Journal of Biological Education, 49*(3), 261–273. https://doi.org/10.1080/00219266.2014.926960

8. Semb, G. B., & Ellis, J. A. (1994). Knowledge taught in school: What is remembered? *Review of Educational Research, 64*(2), 253–286. JSTOR. https://doi.org/10.2307/1170695

9. Wisher, R. A., Curnow, C. K., & Seidel, R. J. (2001). Knowledge retention as a latent outcome measure in distance learning. *American Journal of Distance Education, 15*(3), 20–35. https://doi.org/10.1080/08923640109527091

10. Yielder, J., Bagg, W., & O'Connor, B. (2013). Progress testing: A potential for collaboration and benchmarking across Australian and New Zealand medical schools? *Focus on Health Professional Education, 15*(1), 81–87.