Psychology Learning and Teaching
Volume 10 Number 3 2011
www.wwwords.co.uk/PLAT

Assessment of Differential Learning by Topic in Introductory Psychology

CORINNE L. McNAMARA, ADRIENNE L. WILLIAMSON & TERRENCE D. JORGENSEN
Kennesaw State University, USA

Introductory psychology students typically perform better on posttests compared to pretests; however, not all topics are learned equally well. To measure how much information students learned overall and to determine whether the level of knowledge gained differed by topic, 932 undergraduates enrolled in introductory psychology completed a multiple-choice pretest-posttest assessment. The tests included questions about 11 topics typically taught in introductory psychology. Student-, instructor-, and assessment-related variables that could affect learning were also examined. As expected, posttest scores were significantly higher than pretest scores. Importantly, there were significant differences among topics in terms of posttest scores as well as in the level of improvement. Students were most likely to answer posttest questions correctly about introduction and research methods, memory, and development, and least likely to answer questions correctly about physiological psychology. Scores improved on 9 of the 11 topics, with the greatest improvement for memory, physiological psychology, and sensation and perception. No improvement occurred for cognition and development. Regarding variables that could potentially affect learning, a significant effect for type of final exam, type of credit granted, and instructor was found. This is the first time that a pretest-posttest design has been used to demonstrate differential learning of topics in introductory psychology. Results may inform instructors' course planning regarding time allotted and techniques used to facilitate student learning.

External and internal stakeholders are placing increasing emphasis on assessing learning in academia. External constituencies such as legislators, administrators, and taxpayers require assessment for accountability purposes, whereas professors typically engage in assessment to measure students' acquisition of knowledge, determine teaching effectiveness, and inform pedagogy. There are a variety of ways to assess learning; one commonly used technique is a pretest-posttest measure. This method gauges students' baseline knowledge prior to class and measures acquisition of knowledge over the semester (Linton, 2003).

Several researchers have used pretest-posttest designs to assess learning in introductory psychology (Federici & Schuerger, 1976; Rossi, Keeley, & Buskist, 2005; Thompson & Zamboanga, 2003, 2004). These assessments varied on several characteristics, including number of questions, timing of administration of posttest questions, and similarities between pre- and posttest questions. Rossi et al. (2005) administered a 100-item pretest the first week of class, and a 100-item posttest at the end of the semester, as a cumulative final. Questions were matched for content. Thompson and Zamboanga (2003, 2004) administered a 25-item pretest at the beginning of the semester and integrated six or seven of these questions into each of four noncumulative exams throughout the semester. Not surprisingly, findings from these studies showed that introductory psychology students performed better on posttests than pretests (Federici & Schuerger, 1976; Rossi et al., 2005; Thompson & Zamboanga, 2003, 2004), indicating that students' overall content knowledge increased throughout the course.

 http://dx.doi.org/10.2304/plat.2011.10.3.253

According to past research, some topics in introductory psychology are more difficult to learn than others (Peck, Ali, Matchock, & Levine, 2006). Peck et al. (2006) analysed performance by topic on 968 multiple-choice questions from unit exams of five instructors. They found that questions about social psychology, memory, consciousness, and development were more likely to be answered correctly than questions about physiological psychology and emotion.

There are several plausible explanations for this variability in performance, including student-, instructor-, and assessment-based factors. For example, one student-related variable is major; students who are psychology majors may have more interest in the topics and greater motivation to learn the material than nonmajors. Instructor-related variables that may contribute to this observed difference include instructional style and area of expertise, time allotted to each topic, order of topics covered, textbook selected, and supplemental activities assigned.

Assessment-based factors may also influence variability in performance. Researchers have previously assessed differential learning on unit exams (Peck et al., 2006). However, learning by topic has not been measured using a pretest-posttest design when the assessment is given at the beginning and end of the course. These different approaches introduce variation in the timing of the assessment and the amount of material evaluated at one time. Additionally, whether students are given credit on the assessment may affect motivation, and thus is another variable that could influence performance.

The purpose of our study was twofold: to use a pretest-posttest design to assess introductory psychology students' overall level of learning, and to examine differential learning by topic. Establishing a baseline measurement of prior knowledge and ascertaining the specific topics that are most difficult for students to learn will allow professors mindfully to refine their courses to enhance learning. We also examined student-, instructor-, and assessment-related factors that may contribute to learning.

Method

Participants

Our sample consisted of students enrolled in 24 sections of introductory psychology at a large, comprehensive university in the USA during the fall and spring semesters of one academic year. The pretest was completed by 1518 students, and the posttest was completed by 1128 students. Data analysis was performed on 932 consenting students (632 women, 300 men) who completed both the pretest and posttest. Students enrolled in introductory psychology come from a variety of backgrounds and majors and typically take this class as a foundation course for the major, a general education requirement, or a lower-level elective. Twelve percent of the participants were psychology majors.

Participants ranged in age from 18 to 59 years ($M = 21.51$, $SD = 5.38$) and were primarily European American (76%). Of the remainder, 11% were African American, 2% were Asian American, 6% were Latin American, and 5% identified as other. Forty-six percent of our sample were enrolled in their first year, 30% in their second year, 16% in their third year, and 8% in their fourth year.

Sections of introductory psychology varied on a number of characteristics, including number of students (15-219), duration of class (e.g., 75 minutes twice a week or 150 minutes once a week), time of day, instructor ($N = 15$), and textbook (Myers, 2008 [$n = 4$ instructors]; Weiten, 2008 [$n = 11$]). Instructors generally taught in lecture format and differed in supplemental exercises assigned, time spent lecturing on each topic, order of topics presented, administration of a cumulative or noncumulative final, and type of credit provided for the assessment. Because the score on the assessment was not calculated as part of the student's course grade, some instructors provided either research participation credit ($n = 5$ sections) applied toward fulfilling a course requirement or extra credit ($n = 10$ sections) applied to the total points earned in the course, to encourage students to complete the assessment. Instructors did not provide credit in seven sections (two sections unreported).

Materials

The pretest-posttest assessment was used to evaluate learning on topics from 11 key chapters that were required to be taught in all sections of the course. Chapters were selected by a group of faculty who regularly teach introductory psychology. The topics included introduction and research methods, physiological psychology, nature-nurture and diversity, development, sensation and perception, learning, memory, cognition, personality, psychological disorders and therapy, and social psychology. These topics are somewhat similar to the 16 topics analysed by Peck et al. (2006), except that we combined introduction and research methods as well as sensation and perception, added the topic of nature-nurture and diversity, and did not assess consciousness, motivation, emotion, or health.

Five four-option, multiple-choice questions on each of the 11 topics were selected from Brink's (2008) test bank that accompanies *Exploring Psychology* (Myers, 2008). Five paper-and-pencil versions of an 11-item assessment were created. Each version included one of the five questions on each topic. Each topic was assessed using three to four factual-definitional questions and one to two conceptual-application questions. The pretests and posttests were identical.

Procedure

Students were randomly assigned to complete one of the five versions of the assessment during their regular class time. Individuals completed the same version of the pretest and posttest. Participants completed the pretest during the first (n = 19 sections) or second (n = 5) week of classes. Participants completed the posttest on the day of the final (n = 5), during the last class meeting (n = 12), or within a week of the last meeting (n = 7).

Results

Pretest and posttest scores were calculated as the percentage correct on the 11-item test for each student. A paired-samples t test revealed that the posttest scores (M = 58.27, SD = 18.66) were significantly higher than the pretest scores (M = 49.62, SD = 17.36), $t(931) = 13.80$, $p < .0001$, d = 0.45. We then analysed our dependent variable as a binary outcome in order to assess whether the odds of accuracy differed significantly among topics as well as whether those differences changed from pretest to posttest. Answers for each individual question were coded as 1 (*correct*) or 0 (*incorrect*), using time (pretest or posttest) and question number (1-11) as predictor variables. To control for dependence of observations, we conducted a repeated-measures logistic regression using PROC GENMOD in SAS 9.2.

We found significant differences in odds of accuracy among individual topics on both the pretest, $\chi^2(10, N = 932) = 326.12$, $p < .0001$, and the posttest, $\chi^2(10, N = 932) = 220.27$, $p < .0001$. Table 1 presents topics in the order of decreasing accuracy on the posttest. The interaction between time (pre- and posttest) and question number (i.e., content area) revealed that the degree of improvement from pretest to posttest differed among the 11 questions, $\chi^2(10, N = 932) = 110.00$, $p < .0001$. As seen in Table 2, significant improvements were observed from pretest to posttest on 9 of the 11 topics (i.e., memory, physiological psychology, sensation and perception, personality, learning, psychological disorders and therapy, nature-nurture and diversity, introduction and research methods, and social psychology) whereas no significant change was observed for cognition or development. There was at least a 12-point increase in the percentage of students who answered correctly on items in five areas: memory, physiological psychology, sensation and perception, personality, and learning. Table 2 presents topics in the order of the degree of change from pretest to posttest, along with odds ratios to measure effect size.

We examined several variables that could potentially affect overall and differential learning observed in this study. These variables included student-related (i.e., student major [psychology vs. nonpsychology major]), instructor-related (i.e., order of topic coverage, time allotted to each topic, textbook, supplemental assignments, and instructor), and assessment-related factors (i.e., type of final exam [cumulative or noncumulative], type of credit students received for completing the assessment [research, extra, or no credit], and time of administration of the pre- and posttests). We

added each of these variables to the above model, one at a time, to determine whether they could explain a significant amount of variability in accuracy.

Table 1. Scores ranked by topic on the posttest.

Topic	% Correct	Pretest rank [a]						
Introduction and research methods	70.29	2	A					
Memory	69.65	5	A					
Development	63.64	1		B				
Social psychology	58.93	3			C			
Learning	58.07	6			C			
Personality	57.21	8			C	D		
Sensation and perception	57.21	9			C	D		
Cognition and intelligence	54.64	4			C	D	E	
Nature-nurture and diversity	53.35	7				D	E	
Psychological disorders and therapy	51.38	10					E	
Physiological psychology	46.28	11						F

Note. Topics that share a letter are not significantly different.
[a] Pretest scores are available in Table 2.

Table 2. Ranked change in scores between pretest and posttest.

Topic	Pretest	Posttest	% Change	Z	p	OR[a]						
Memory	52.28	69.65	17.37	7.94	< .0001	2.09	A					
Physiological psychology	29.56	46.28	16.72	7.68	< .0001	2.05	A	B				
Sensation and perception	44.35	57.21	12.86	5.70	< .0001	1.68	A	B	C			
Personality	44.78	57.21	12.43	5.50	< .0001	1.65		B	C			
Learning	46.06	58.07	12.00	5.31	< .0001	1.62			C	D		
Psychological disorders and therapy	43.39	51.38	7.99	3.52	.0004	1.38			C	D	E	
Nature-nurture and diversity	46.17	53.35	7.18	3.16	.0016	1.33				D	E	
Introduction and research methods	64.61	70.29	5.68	2.67	.0076	1.30					E	F
Social psychology	54.42	58.93	4.50	2.00	.0458	1.20					E	F
Cognition	52.28	54.64	2.36	1.04	.2992	1.10						F
Development	67.61	63.64	-3.97	-1.84	.0666	0.84						G

Note. Topics that share a letter are not significantly different regarding percent change.
[a] Odds ratios (OR) compare posttest odds of accuracy to pretest odds of accuracy - e.g., for personality questions, posttest odds of accuracy are 1.65 times (or 65% higher than) the pretest odds of accuracy.

There were no significant differences in overall posttest scores between psychology majors and nonmajors, $\chi^2(1) = 0.53$, $p = .47$, $N = 932$. A significant effect of instructor on posttest scores by topic was also observed, $\chi^2(14) = 35.10$, $p = .001$, $N = 932$. None of the other instructor-related variables had a significant effect on performance ($p > .10$).

Final exam type and credit type did not affect posttest accuracy differentially by topic, but these factors did have a significant effect on overall posttest scores; however, neither effect was in the expected direction. A 2 (final exam type: cumulative or noncumulative) × 3 (credit type: research, extra, or none) ANOVA on posttest scores revealed a main effect of exam type, such that students whose final exam was cumulative ($M = 57.11$, $SD = 18.96$) scored lower than students

whose final exam was noncumulative ($M = 60.29$, $SD = 17.97$), $F(1, 885) = 5.53$, $p = .02$, partial $\eta^2 = .01$. There was also a main effect of credit type, $F(2, 885) = 4.43$, $p = .01$, partial $\eta^2 = .01$. Tukey's HSD posthoc test revealed that students who received no credit ($M = 61.05$, $SD = 18.23$) for completing the assessment scored similarly to students who received extra credit ($M = 57.86$, $SD = 19.03$) but significantly higher than students who received research credit ($M = 56.48$, $SD = 18.79$); there was no significant difference between extra credit and research credit. There was no significant interaction between final exam type and credit type, $p > .99$. Additionally, there was no significant difference for the timing of pretest or posttest administration (i.e., when students completed the pretest relative to the first lecture or when students completed the posttest relative to the final exam), $p > .10$.

Discussion

As expected, posttest scores were significantly higher than pretest scores on the course assessment, with a moderate effect size. These results are consistent with previous research that has shown improvement in knowledge in introductory psychology during the semester (Federici & Schuerger, 1976; Rossi et al., 2005; Thompson & Zamboanga, 2003, 2004). Our pretest score of 49.6% was slightly above the range of previous pretest scores, which varied from 17.9% to 44.4% (Federici & Schuerger, 1976; Thompson & Zamboanga, 2003, 2004). This discrepancy may be due to timing of the pretest. In prior studies, researchers administered the pretest within the first week of class whereas 5 of our 24 sections did not complete the pretest until the second week of class. Covering material before the pretest may have inflated our students' average on the pretest; however, there were no significant differences in the pretest scores between students who completed the pretest during the first and second weeks of class.

Our posttest score of 58.3% was lower than other posttest scores, which varied from 76.4% to 83.4% (Federici & Schuerger, 1976; Thompson & Zamboanga, 2003, 2004). Our lower posttest scores may be due to two important factors: motivation and level of preparation for the posttest. In prior studies, students' performance on posttest questions was calculated as part of their course grade, therefore students were motivated to do well on the test and had likely studied shortly before the exam. In our study, students' performance was not calculated as part of their grade. Consequently, it is possible that their level of motivation was lower, and only the most recent material had been studied shortly before the posttest assessment was administered.

Importantly, there were significant differences in posttest scores among topics, as well as significant variability in levels of improvement from pretest to posttest among topics, showing that students learn some topics better than others. On the posttest, students earned passing scores (i.e., greater than 60%) on introduction and research methods, memory, and development. Students showed significant improvement in test scores on 9 of the 11 topics, with the greatest improvement on memory, physiological psychology, and sensation and perception. Similarly, Peck et al. (2006) found that students were more likely to answer questions correctly about memory compared to most other topics, indicating that memory appears to be one of the better learned topics in introductory psychology. Social psychology was the highest scoring topic for students in Peck et al.'s study, with 85% of the questions being answered correctly. In our study, posttest scores on social psychology were ranked fourth out of 11, but students' average of 58.9% did not demonstrate a passing level of knowledge.

Our students demonstrated the least amount of knowledge about physiological psychology, although this concept was second in terms of improvement, with an increase of 16.72 points. Peck et al. (2006) also found that students performed lowest on physiological psychology. The only topics for which students did not demonstrate a significant pretest-posttest difference were cognition and development, although it is unclear why this would be the case. Scores on development might not have improved because they were the highest pretest scores, leaving less room for improvement, but the same cannot be said for scores on cognition. Researchers should examine why students failed to improve on these topics.

Time between study and test may explain differences in findings between our study and that of Peck et al. (2006). Peck and colleagues assessed performance on questions from unit exams, so it is likely that students studied the material shortly before taking these exams. We administered our

posttest at the end of the semester, so students had not studied some of the material since earlier in the semester.

The only variables that we found to significantly affect overall posttest scores were type of final exam and type of credit granted. One might expect that studying all topics for a cumulative exam might yield higher overall scores on the posttest; however, in the majority of the sections that had a cumulative exam, the posttest was administered prior to the day of the final, so students may not have reviewed material before the posttest. Furthermore, there was no difference in scores based on timing of posttest administration within these sections. Unexpectedly, we found that students who completed a cumulative exam actually scored lower on the posttest. Future research will be necessary to further explore this effect.

We did not expect to find a difference on scores based on type of credit given because credit was granted for completion of the assessment, rather than for accuracy of their answers; however, we found that students who did not receive credit performed better on the posttest than those who received research credit. It may be that students who were given research credit were extrinsically motivated, whereas students who were not granted credit were intrinsically motivated to accurately complete the assessment, but this explanation does not account for why we found no difference between students who received extra credit and those who did not receive credit.

A significant effect was observed for instructor on posttest scores by topic. It is difficult to determine the factors that explain this finding. Because type of final exam and type of credit granted differed by instructor, it is possible that the effect of instructor is merely an artifact of these factors. Furthermore, there are other variables that we did not measure that may explain the effect of instructor, such as their experience, expertise, and areas of interest. Future research is necessary to examine the effects of student-, instructor-, and assessment-related variables on learning.

Limitations and Future Research

There are some limitations that should be considered when interpreting our results. First, the list of topics that we assessed was broad and based on chapter titles. Second, instructors were not required to cover specific concepts within a chapter. Instructors may have emphasised different concepts within a topic based on their areas of specialisation and interests; therefore, it is possible that we assessed students' knowledge of a concept that was not explicated in their class. Finally, for six sections, instructors provided supplemental assignments that may have accentuated differential learning. Because detailed information about those assignments was not available, we could not assess the effect of this potential confound by topic; however, we found no effect of supplemental assignments on overall performance.

Despite these limitations, this research is important for introductory psychology instructors. To the best of our knowledge, this is the only study implementing a pretest-posttest design to show that students know significantly more about some topics before taking the class and that some topics are easier to learn than others. We also examined the effects of student-, instructor-, and assessment-related variables on overall and differential learning. Because instructors must determine the breadth and depth of topics to cover within a semester, knowing which topics students struggle with and what factors influence learning may inform instructors how to allocate their class time.

We plan to revise our pretest-posttest measure to directly align each assessment item with a specific learning objective for introductory psychology. Additionally, we are developing activities on topics that students find most challenging. For example, although students' scores significantly improved on physiological psychology, their posttest scores remained the worst of all the topics. Therefore, we plan to have students create a concept map, an effective technique for better understanding connections between key concepts (Nesbit & Adesope, 2006). Research indicates that there are various additional approaches for improving learning, such as generating topic-relevant questions (Berry & Chew, 2008), writing assignments (Drabick, Weisberg, Paul, & Bubier, 2007; Stewart, Myers, & Culley, 2010), and in-class self-assessment using clickers (Morling, McAuliffe, Cohen, & DiLorenzo, 2008; Shaffer & Collura, 2009). Other departments should develop similar assessments of their introductory psychology students to reveal specific

deficiencies, unique to that department, which might benefit from more attention, and develop appropriate assignments to improve student performance.

Note

Corinne McNamara and Adrienne Williamson contributed equally to this article and are listed in alphabetical order for convenience purposes only.

References

Berry, J. W., & Chew, S. L. (2008). Improving learning through interventions of student-generated questions and concepts maps. *Teaching of Psychology, 35,* 305-312. http://dx.doi.org/10.1080/00986280802373841

Brink, J. (2008). *Test bank to accompany David G. Myers'* Exploring psychology (7th ed.). New York: Worth Publishers.

Drabick, D. A. G., Weisberg, R., Paul, L., & Bubier, J. L. (2007). Keeping it short and sweet: Brief, ungraded writing assignments facilitate learning. *Teaching of Psychology, 34,* 172-176. http://dx.doi.org/10.1080/00986280701498558

Federici, L., & Schuerger, J. (1976). High school psychology students versus non-high school psychology students in a college introductory class. *Teaching of Psychology, 3,* 172-174. http://dx.doi.org/10.1207/s15328023top0304_4

Linton, D. (Ed.) (2003). Focus on assessment. *First II: Faculty Institutes for Reforming Science Technology, 1*, 1-8.

Morling, B., McAuliffe, M., Cohen, L., & DiLorenzo, T. M. (2008). Efficacy of personal response systems 'clickers' in large, introductory psychology classes. *Teaching of Psychology, 35,* 45-50. http://dx.doi.org/doi:10.1080/00986280701818516

Myers, D. (2008). *Exploring psychology* (7th ed.). New York: Worth Publishers.

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76,* 413-448. http://dx.doi.org/10.3102/00346543076003413

Peck, A. C., Ali, R. S., Matchock, R. L., & Levine, M. E. (2006). Introductory psychology topics and student performance: Where's the challenge? *Teaching of Psychology, 33,* 167-170. http://dx.doi.org/doi:10.1207/s15328023top3303

Rossi, M., Keeley, J., & Buskist, W. (2005). High school psychology performance in the college introductory psychology course. *Teaching of Psychology, 32,* 52-54.

Shaffer, D. M., & Collura, M. J. (2009). Evaluating the effectiveness of a personal response system in the classroom. *Teaching of Psychology, 36,* 273-277. http://dx.doi.org/doi:10.1080/00986280903175749

Stewart, T. L., Myers, A. C., & Culley, M. R. (2010). Enhanced learning and retention through 'writing to learn' in the psychology classroom. *Teaching of Psychology, 37,* 46-49. http://dx.doi.org/doi:10.1080/00986280903425813

Thompson, R. A., & Zamboanga, B. L. (2003). Prior knowledge and its relevance to student achievement in introduction to psychology. *Teaching of Psychology, 30,* 96-101. http://dx.doi.org/oi:10.1207/s15328023top3002

Thompson, R. A., & Zamboanga, B. L. (2004). Academic aptitude and prior knowledge as predictors of student achievement in introduction to psychology. *Journal of Educational Psychology, 96,* 778-784. http://dx.doi.org/10.1037/0022-0663.96.4.778

Weiten, W. (2008). *Psychology: Themes and variations* (8th ed.). Belmont, CA: Wadsworth.

CORINNE McNAMARA* has been an assistant professor of psychology at Kennesaw State University since 2006. She earned her BA in psychology from the University of Mississippi. After earning her MA in neuroscience from Baylor University, she returned to the University of Mississippi, where she earned her PhD in experimental psychology. Dr McNamara regularly teaches introductory psychology, research methods, and physiological psychology. Her primary research interests are in the areas of interpersonal violence and the scholarship of teaching and learning. *Correspondence*: Corinne McNamara, Kennesaw State University, 1000 Chastain Road, MD 2202, Social Science Building 22, Kennesaw, GA 30144, USA (cmcnama4@kennesaw.edu).

ADRIENNE WILLIAMSON has been an assistant professor of psychology at Kennesaw State University since 2006. She earned her MS in psychology from Augusta State University and her PhD in experimental psychology with a subspecialty in neuropsychology from the University of Memphis. She has recently taught experimental psychology, physiological psychology, psychopharmacology, and introductory psychology. Her research focuses on determining factors that influence student learning and metacomprehension. Additionally, Dr Williamson is interested in examining cognitive sequelae of neurological disorders. *Correspondence*: awill176@kennesaw.edu

TERRENCE JORGENSEN earned his BS in psychology and MS in applied statistics from Kennesaw State University. He currently works as a research tutor in the KSU psychology lab, for undergraduates learning research methods, statistical analysis, and APA style. His teaching interests include categorical and longitudinal modelling, on which he is focusing while earning his PhD in quantitative psychology at the University of Kansas. *Correspondence*: tdj4728@kennesaw.edu

*Contact author

Manuscript submitted 6 January 2011
Revision accepted for publication 28 March 2011