International Journal of Engineering Education Vol. 35, No. 4, pp. 1170–1175, 2019
Printed in Great Britain
0949-149X/91 $3.00+0.00

Exploring the Role of Testing in Student Outcomes: Evidence from a Mechanics Course*

PAUL BAZELAIS[1], DAVID JOHN LEMAY[1] and TENZIN DOLECK[2]
[1] McGill University, 3700 McTavish Street, Montreal, Quebec, H3A 1Y2, Canada. E-mail: paul.bazelais@mail.mcgill.ca; david.lemay@mail.mcgill.ca
[2] University of Southern California, 3470 Trousdale Pkwy, Los Angeles, CA 90089, USA. E-mail: doleck@usc.edu

Assessments have become increasingly prevalent in education. While many affordances of assessments are offered in the literature, there is mixed evidence on how assessments affect students' learning and performance. Moreover, a testing effect has been identified in lab-based studies where more testing is associated with better performance; however, less is known about the effects of testing on performance in situ. The present study employs data from two Mechanics courses to analyze the effects of testing on performance. We compare two sections—experimental condition with testing (N = 36) and control condition with homework (N = 38)—of the Mechanics course, to examine the relative importance of testing. We find a strong effect for regular testing on student mid-term and final exam performance. The findings have broad implications for the growing testing effect literature.

Keywords: testing effect; learning outcomes; STEM education; pre-university; physics; academic performance

1. Introduction

Improving students' learning outcomes is an institutional mandate of educational organizations. Learning outcomes are generally measured through the use of some form of assessment [1]. Today, students regularly face some form of assessment in their academic lives. In fact, in the present age of accountability management, many organizations are impelled to collect extensive performance data and students regularly sit for high-stakes examinations. Indeed, assessment has become for better or worse a distinctive and ubiquitous feature of education [2–4]. Among a wide range of assessments, testing is the primary means to evaluate learning and achievement [5] as "[it] is a form of assessment that occurs in the classroom" [6, p. 223]

Many researchers have sought to better understand the effects of testing [7–17]. One area of research has become focused on explicating the *testing effect*, where "students who take a test on material between the time they first study and the time they take a final test remember more of the material than students who do not take an intervening test" [10, p. 392]. Examining the effects of testing is especially important because of the implications on students' learning and performance outcomes [15, 16, 18] and the prevalence of testing in today's education [1, 19, 20].

The testing effect has been demonstrated with varied test formats and study materials and in different educational settings [20–22]. A recent meta-analysis has shown that for learning, practice tests are more beneficial than restudying [19]. Beyond improved retention of studied material over restudying [15, 22], research has also demonstrated a link between testing and skills learning [23]. In fact, "taking a test can do more than simply assess learning: tests can also enhance learning and improve long-term retention" [24, p. 861].

Thus far, the growing literature that empirically investigates the testing effect has primarily focused and relied on experimental studies; Butler and Roediger [5] examined the testing effect in a simulated classroom setting; Agarwal, Karpicke, Kang, Roediger, and McDermott [24] conducted an experimental study of the testing effect where students were tested on prose passages with open-book and closed-book tests; Carpenter, Pashler, Wixted, and Vul [25] showed experimentally that testing (via memory tests) enhanced overall recall more compared to restudying; Verkoeijen, Bouwmeester, and Camp [26] observed the testing effect in an experimental comparison of learning wordlists through either restudying or testing.

Despite the importance of evaluating the effects of testing, there is comparatively less research on the testing effect in natural settings. Hence, little is known about the degree to which we can generalize from controlled lab experimental studies to real-world learning situations. Moreover, there has been some debate regarding the testing effect owing to the mixed evidence [27–29] . If testing potentially can influence students' learning and performance, then additional in situ examinations of the connections between testing and student outcomes are warranted that go beyond the controlled setting of experimental studies.

* Accepted 17 April 2019.

Table 1. Summary of Methodology

Sections	Condition	Outcome Measures
Treatment group	Lecture format with testing effect	Quizzes, unit tests, standardized final exam (FX)
Control group	Lecture format with online homework	Online homework, unit tests, standardized final exam (FX)

Note. Each section was taught by a different instructor.

Where previous studies considered the testing effect in lab-based and experimental studies, the present study responds to recent calls for more investigations of the testing effect in natural educational contexts [14, 21, 30] and focuses on testing in natural learning settings by examining how testing affects students' academic outcomes in a physics course. Specifically, we compare data from pre-university science students enrolled in two Mechanics courses to understand the effect of testing on academic outcomes.

Table 2. The Sample

Sample		Treatment Group	Control Group
N		36	38
Gender	Male	44%	58%
	Female	56%	42%
HSA		85.39%	83.53%

Note. A one-way ANOVA analysis shows no significant differences, $p > 0.05$.

2. Methodology

2.1 Research context

The current study involves a comparative case study that contrasts two sections of a pre-university Mechanics course. A treatment group included a course that used robust testing with no assigned homework and a control group included a course with online homework and instant formative feedback. To eliminate teacher bias from a comparative case study, the two sections were taught by different instructors with identical content, including three required unit tests (e.g., test 1 on week 5, test 2 on week 10, and test 3 on week 15) and a standardized final exam. In addition, the three required unit tests and the standardized final exam were identical for both the sections. Table 1 illustrates the two conditions in the present study. For each condition, the outcome measures (such as quizzes, unit tests, homework, and standardized final exam (FX) are also listed.

Compared to the control group, who had assigned weekly homework, the students in the treatment group were quizzed 12 times during the 15-week semester with no additional homework. The quizzes were designed with the intent to promote better learning outcomes and provide formative peer feedback in preparation for the three required tests and the final exam. Students were given fifteen minutes to complete individually each quiz. Thereafter, students were given an opportunity (ten minutes) to discuss the quiz with their peers as a way to get formative feedback.

Before the study was conducted, participants were informed of the voluntary and confidential nature of the study. Students who consented to participate were assured that study results would not be linked to any identifiable data. The data included in the current study was not analyzed until after the final grades were submitted.

2.2 Study participants

The current study was conducted using data from first-semester college physics students at an English *Collège d'enseignement général et professionnel* (CEGEP) in Montreal, Quebec (for a primer on CEGEPs, see [31]). The sample ($N = 74$, 51% males, 49% females) was drawn from two sections of the Mechanics Physics course. The treatment group consisted of $N = 36$ students (56% females, 44% males) and the control group consisted of $N = 38$ students (42% females, 58% males).

In order to rule out systematic bias, comparative statistics were used to analyze the sample. No systematic differences between the two groups were found (as illustrated in Table 2). The High School Average (HSA) was essentially the same for the control group ($N = 38$, HSA = 83.53%, Std. Deviation = 4.49) and the treatment group ($N = 36$, HSA = 85.39%, Std. Deviation = 4.40). A one-way ANOVA analysis shows that these two groups were not significantly different, $F(1,70) = 3.33$, $p > 0.05$), at the beginning of the semester. In addition, there was no significant differences between the genders, $F(1,70) = 0.313$, $p > 0.05$.

3. Analysis and results

The data were analyzed in the spirit of Zhang, Ding, and Mazur [32] using within-sample paired t-tests to assess if there was any significant shift within and between the two groups (treatment and control) on the unit tests average ($M = 73.36\%$, $SD = 13.41$) and the final exam average ($M = 66.33\%$, $SD = 15.43$).

Table 3. Overall shift between unit tests and final exam scores for each group

Sections	Tests avg % (*SD*)	FX avg. % (*SD*)	Shift % (*SD*)	*t*	t-test *p*	*ES*
Treatment group	74.00 (14.80)	70.40 (15.18)	3.60 (8.33)	2.60	0.014	0.43
Control group	72.76 (12.12)	62.49 (14.85)	10.27 (19.93)	3.18	0.003	0.51
Both groups	73.36 (13.41)	66.33 (15.43)	7.03 (15.68)	3.86	0.001	0.45

Table 4. Overall difference between the treatment and the control group

Sections	Unit tests avg. % (*SD*)	FX avg. % (*SD*)	Shift % (*SD*)
Treatment group	74.00 (14.80)	70.40 (15.18)	3.60 (8.33)
Control group	72.76 (12.12)	62.49 (14.85)	10.27 (19.93)
t-test	0.397	**2.26***	–1.86

*Note. * $p < 0.05$.*

The overall shift between the unit tests average and the final exam average was statistically significant for the two groups combined, [t (73) = 3.86, p < 0.001, ES = 0.45], whereas, the shift difference for the treatment group was also statistically significant (see Table 3), revealing that students in the treatment group performed better by 3.60 points in the three unit tests average as compared to their performance in the standardized final exam, [t (35) = 2.60, p = 0.014, ES = 0.43].

In addition, a within-sample paired *t*-test found a strong significant positive shift for the control group, [t (37) = 3.18, p = 0.003, ES = 0.51], revealing that students in the control group were less successful on the standardized final exam by over 10 points (see Table 2). Using Cohen's d [33] criterion of significance if $d > 0.2$, effect size (ES) demonstrated that the observed changes were both of statistical and practical significance. The within-sample paired *t*-test difference between the treatment and the control group shows that robust testing can lead to better later retention and performance in a standardized final exam in a college science course.

The overall unit tests average (see Table 4) for the treatment group (M = 74.00%, SD = 14.80) was slightly higher compared to the control group (M = 72.76%, SD = 12.12). However, an independent *t*-test revealed that this difference was not statistically significant, [t (72) = 0.397, p = 0.693, ES = 0.092]. In addition, an independent *t*-test demonstrated that both the overall standardized final exam average difference between the treatment group (M = 70.40%, SD = 15.18) and the control group (M = 62.49%, SD = 14.85) was statistically significant, [t (72) = 2.26, p = 0.027, ES = 0.53] (see Table 4), providing preliminary evidence that robust testing and peer formative feedback may enhance later retention and performance.

3.1 Gender differences

An independent t-test—to evaluate the overall effects (both groups combined) of achievement gains, overall academic performance and HSA by gender—revealed that both males and females do not exhibit any change (see Table 5). Furthermore, the independent t-tests did not reveal any overall difference across all examined variables, suggesting no differences across gender before and after the standardized final exam.

Gender differences within each group on achievement gain, pre and post final exam were also assessed using an independent *t*-test. For the control group, no significant differences were observed between genders across all variables pre and post final exam (see Table 6).

Similarly, the independent *t*-test revealed no significant gender differences in the treatment group across all the examined variables (see Table 7). It is interesting to note that, despite the nontrivial differences between the genders overall, female students recorded higher final exam average (M = 71.31%, SD = 17.58) compared to their male counterparts

Table 5. Overall HSA, unit tests and final exam average, and shift by gender for both groups

Gender	Unit test avg. % (*SD*)	FX avg. % (*SD*)	Shift % (*SD*)	HSA % (*SD*)
Males (n = 38)	72.91 (12.78)	65.23 (14.98)	7.68 (18.40)	84.58 (4.95)
Females (n = 36)	73.84 (14.21)	67.50 (16.02)	6.34 (12.40)	84.28 (4.07)
t-test results	p = 0.769	p = 0.531	p = 0.716	p = 0.777

Note. An independent t-test was not significant across all variables.

Table 6. Average unit tests, final exam, and shift for the control group

Gender	Unit test avg. % (*SD*)	FX avg. % (*SD*)	Shift % (*SD*)
Males (n = 22)	72.18 (13.72)	62.31 (16.47)	9.87 (23.21)
Females (n = 16)	73.55 (9.87)	62.73 (12.82)	10.81 (14.98)
t-test results	p = 0.738	p = 0.932	p = 0.889

Note. No significant differences exist between genders across all variables $p > 0.05$.

Table 7. Average unit tests, final exam, and shift by gender for the treatment group

Gender	Unit tests avg. % (*SD*)	FX avg. % (*SD*)	Shift (%) (*SD*)
Males ($n = 16$)	73.91 (11.73)	69.25 (11.98)	4.66 (7.98)
Females ($n = 20$)	74.07 (17.17)	71.31 (17.58)	2.76 (8.71)
t-test results	p = 0.975	p = 0.691	p = 0.504

Note. An independent t-test analysis shows no significant differences between the genders, $p > 0.05$.

(M = 69.25%, SD = 11.98) in the treatment group. Male students had a higher achievement deficit between the unit test average and the final exam (M = 4.66%, SD = 7.98) compared to female students (M = 2.76%, SD = 8.71); female students scored an average of 2.06 points higher in the standardized final exam than male students. However, this difference was trivial and non-significant, suggesting that these two groups were similar and benefited equally from the testing effect and peer feedback.

4. Discussion

Students appear to self-regulate their performance based on formative feedback [34]. It appears clear that testing condition strongly improved student performance, one may reasonably conclude that testing leads to improved performance and therefore, learning. However, as content was explicitly taught based on the learning objectives and competences, the improved performance may be an artifact of rehearsal rather than learning. Critics of teaching to the test [35] argue that it is no measure of learning but rather speaks to the benefits of rehearsal on test performance. Indeed, test performance is only a proxy of learning and a relatively poor one as longitudinal studies employing test-retest methodologies months or years later report poorer performances after delay except where knowledge and skills continued to be practiced.

Both groups had the opportunity to practice in the three unit tests before the final exam. The only difference between the two groups was: the treatment group had quizzes with peer formative feedback and the control group had weekly online homework with instant feedback. Both groups had a chance to practice with formative feedback, except for the treatment group there were no outside class assignments. Compared to the testing condition, homework-only students had the chance to expand their understanding of the class lectures through the homework assignments, but they did not have the opportunity to discuss their understanding with peers post hoc. Depending on the grading of tests and assignments, there can be different incentive for study between the two conditions. There may be different perceptions and different motivations of the value and objective for studying; homework can be considered an assessment for learning, while a test is an assessment of learning [36]. It is likely that the homework and testing instructional designs induced different study approaches to the material [37]. Thus, the testing condition might induce a surface approach to studying oriented towards performance and away from deep understanding (Biggs, 1987). More longitudinal of approaches to studying have demonstrated better gains over the long-term for deeper compared to surface learning approaches [38].

Much, as in a previous study [30], we found an immediate effect for formative assessment that suggests that students respond to formative feedback by adjusting and regulating their performance. It appears likely that aspects of the instructional situation may have a confounding effect on the present findings. The test-taking treatment condition also included a peer review and exchange activity that provided feedback that was "nonevaluative, supportive, timely, and specific" [39, p. 153], that is, feedback that is tuned to be maximally effective for learning [39]. In comparison, the control group did not benefit from a peer review and feedback session. Thus, it remains unclear whether it is the frequent testing or the supporting activities that are related to improved performance in Mechanics.

The present conversation on the testing effect recalls the debate on mastery and meaningful learning [40]. Mastery learning or teaching to curriculum objectives offered structured curricula and clear, measurable outcomes. However, it was a piecemeal vision of learning and offered a limited vision of education. Mastery learning, embodying a behaviorist epistemology, rendered learning as the accumulation of discrete bits of information and knowledge was cast as declarative, procedural, and strategic. Yet, critics argued from a constructivist perspective that this limited model of learning did not address the black box of cognition and more contextual forms of knowledge and knowing grounded in social activity; indeed, that the learner was an active participant in the construction of knowledge. Teaching to the test may present learning gains in the short term however over the long-term these prove rather illusory. Indeed, such rote learning limits the agency of the individual and knowledge thus gained remains limited and unintegrated, dissociated from meaningful experience.

To advance the debate over the testing effect, more research ought to be conducted in naturalistic settings that take a more multifaceted and multifactorial approach to the study of teaching and learning [41]. Studies of the testing effect need to consider a wider range of learning outcomes including repeated measures, but also attitudinal measures and measures of learning transfer. To that end, mixed-methods approaches [42] appear to have the most potential for identifying salient factors and processes prevailing across social, cognitive, and affective dimensions of the learning environment.

4.1 Implications

On the surface, our study appears to confirm the testing effect as students in the repeated testing group outperformed students in the homework-only control group on mid-term and final exams. However, our study also raises the possibilities of confounding variables inherent in the operationalizations of the treatment and control which induced uncontrollable variations between the groups. As a rare case study of the testing effect conducted in a naturalistic setting, this research highlights the difficulties of attributing effects solely to the treatment condition and stresses the need for further more complex methodologies that can capture the multifactorial classroom reality to understand the range of interactions—both affordances and constraints [38]—which support effective classroom learning environments.

4.2 Limitations

We did not collect any participant profile and cannot speculate on how the testing condition influences learners' behaviors, including goals, motivation, and approach to learning. The present study is limited by its use of a non-randomized convenience sample. However, using two sections in the same semester taught using the same materials mitigated bias across groups. This is further supported by the absence of any significant effects for gender across both sections. The cross-sectional nature of the study limits the conclusions about the persistence of knowledge gains over the long-term.

4.3 Future directions

The present study can be extended by examining how frequent testing influences learners' across cognitive, social, and affective dimensions to understand the influence of regular testing on perceptions of the learning environment and to better understand the relationship between testing and learning, not simply in terms of testing performance, but in terms of lasting, long-term learning gains.

5. Conclusion

We find a strong testing effect for two Mechanics sections, comparing a testing condition with a control group. We do not find any differences between males and females. The findings reported in the present study have broad implications for the growing testing effect literature.

References

1. S. Binks, Testing enhances learning: A review of the literature, *Journal of Professional Nursing*, **34**(3), pp. 205–210, 2018.
2. D. Dixson and F. Worrell, Formative and Summative Assessment in the Classroom, *Theory Into Practice*, **55**(2), pp. 153–159, 2016.
3. W. Harlen and M. James, Assessment and Learning: differences and relationships between formative and summative assessment, *Assessment in Education: Principles, Policy & Practice*, **4**(3), pp. 365–379, 1997.
4. A. Lau, 'Formative good, summative bad?'—A review of the dichotomy in assessment literature, *Journal of Further and Higher Education*, **40**(4), pp. 509–525, 2015.
5. A. Butler and H. Roediger, Testing improves long-term retention in a simulated classroom setting, *European Journal of Cognitive Psychology*, **19**(4–5), pp. 514–527, 2007.
6. A. B. Champagne and V. L. Kouba, Writing to inquire: Written products as performance measures, in J. J. Mintzes, J. H. Wandersee and J. D. Novak (eds), *Assessing science understanding: A human constructivist view*, Elsevier, NYC, NY, pp. 223–247, 2005.
7. C. Bae, D. Therriault and J. Redifer, Investigating the testing effect: Retrieval as a characteristic of effective study strategies, *Learning and Instruction*, **60**, pp. 206–214, 2018.
8. P. Delaney, P. Verkoeijen and A. Spirgel, Spacing and Testing Effects, *Psychology of Learning and Motivation*, **53**, pp. 63–147, 2010.
9. L. Eglington and S. Kang, Retrieval Practice Benefits Deduc-

tive Inference, *Educational Psychology Review*, **30**(1), pp. 215–228, 2016.
10. J. Glover, The ''testing'' phenomenon: Not gone but nearly forgotten, *Journal of Educational Psychology*, **81**(3), pp. 392–399, 1989.
11. C. Johnson and R. Mayer, A testing effect with multimedia learning, *Journal of Educational Psychology*, **101**(3), pp. 621–629, 2009.
12. J. Karpicke, A. Butler and H. Roediger III, Metacognitive strategies in student learning: Do students practice retrieval when they study on their own?, *Memory*, **17**(4), pp. 471–479, 2009.
13. T. Kuo and E. Hirshman, Investigations of the Testing Effect, *The American Journal of Psychology*, **109**(3), p. 451, 1996.
14. M. McDaniel, J. Anderson, M. Derbish and N. Morrisette, Testing the testing effect in the classroom, *European Journal of Cognitive Psychology*, **19**(4–5), pp. 494–513, 2007.
15. H. Roediger and J. Karpicke, Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention, *Psychological Science*, **17**(3), pp. 249–255, 2006a.
16. H. Roediger and J. Karpicke, The Power of Testing Memory: Basic Research and Implications for Educational Practice, *Perspectives on Psychological Science*, **1**(3), pp. 181–210, 2006b.
17. Á. Szöllösi, A. Keresztes, B. Novák, B. Szászi, S. Kéri and Racsmány, The Testing Effect is Preserved in Stressful Final Testing Environment, *Applied Cognitive Psychology*, **31**(6), pp. 615–622, 2017.
18. R. Bangert-Drowns, J. Kulik and C. Kulik, Effects of Frequent Classroom Testing, The *Journal of Educational Research*, **85**(2), pp. 89–99, 1991.
19. O. Adesope, D. Trevisan and N. Sundararajan, Rethinking the Use of Tests: A Meta-Analysis of Practice Testing, *Review of Educational Research*, **87**(3), pp. 659–701, 2017.
20. H. Roediger and J. Karpicke, Reflections on the Resurgence of Interest in the Testing Effect, *Perspectives On Psychological Science*, **13**(2), pp. 236–241, 2018.
21. W. Batsell, J. Perry, E. Hanley and A. Hostetter, Ecological Validity of the Testing Effect, *Teaching of Psychology*, **44**(1), pp. 18–23, 2016.
22. S. Pan and T. Rickard, Transfer of test-enhanced learning: Meta-analytic review and synthesis, *Psychological Bulletin*, **144**(7), pp. 710–756, 2018.
23. C. Kromann, M. Jensen and C. Ringsted, The effect of testing on skills learning, *Medical Education*, **43**(1), pp. 21–27, 2009.
24. P. Agarwal, J. Karpicke, S. Kang, H. Roediger and K. McDermott, Examining the testing effect with open- and closed-book tests, *Applied Cognitive Psychology*, **22**(7), pp. 861–876, 2008.
25. S. Carpenter, H. Pashler, J. Wixted and E. Vul, The effects of tests on learning and forgetting, *Memory & Cognition*, **36**(2), pp. 438–448, 2008.
26. P. Verkoeijen, S. Bouwmeester and G. Camp, A Short-Term Testing Effect in Cross-Language Recognition, *Psychological Science*, **23**(6), pp. 567–571, 2012.
27. J. Karpicke and W. Aue, The Testing Effect Is Alive and Well with Complex Materials, *Educational Psychology Review*, **27**(2), pp. 317–326, 2018.
28. D. Peterson and N. Mulligan, The negative testing effect and multifactor account, *Journal of Experimental Psychology: Learning, Memory, And Cognition*, **39**(4), pp. 1287–1293, 2013.
29. T. van Gog and J. Sweller, Not New, but Nearly Forgotten: the Testing Effect Decreases or even Disappears as the Complexity of Learning Materials Increases, *Educational Psychology Review*, **27**(2), pp. 247–264, 2015.
30. P. Bazelais, T. Doleck and D. J. Lemay, Exploring the Association Between Formative and Summative Assessment in a Pre-University Science Program, *Journal of Formative Design in Learning*, **1**(2), pp. 65–72, 2017.
31. P. Bazelais, D. J. Lemay and T. Doleck, How Does Grit Impact Students' Academic Achievement in Science?, *European Journal of Science and Mathematics Education*, **4**(1), pp. 33–43, 2016.
32. P. Zhang, P. L. Ding and E. Mazur, Peer Instruction in introductory physics: A method to bring about positive changes in students' attitudes and beliefs, *Physical Review Physics Education Research*, **13(**1), pp. 1–10, 2017.
33. J. Cohen, A power primer, *Psychological Bulletin*, **112**, pp. 155–159, 1992.
34. D. J. Nicol and D. Macfarlan-Dick, Formative assessment and self-regulated learning: a model and seven principles of good feedback practice, *Studies in Higher Education*, **31**(2), pp. 199–218, 2006.
35. R. E. Bennett, Formative assessment: a critical review, *Assessment in Education: Principles, Policy & Practice*, **18**(1), pp. 5–25, 2011.
36. W. Harlen and M. James, Assessment and Learning: differences and relationships between formative and summative assessment, *Assessment In Education: Principles, Policy & Practice*, **4**(3), pp. 365–379, 1997.
37. J. Biggs, Enhancing teaching through constructive alignment, *Higher Education*, **32**(3), pp. 347–364, 1996.
38. M. Baeten, E. Kyndt, K. Struyven and F. Dochy, Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness, *Educational Research Review*, **5**(3), pp. 243–260, 2010.
39. V. Shute, Focus on Formative Feedback, *Review of Educational Research*, **78**(1), pp. 153–189, 2008.
40. R. E. Mayer, Rote Versus Meaningful Learning, *Theory into Practice*, **41**(4), pp. 226–232, 2002.
41. R. Kane, S. Sandretto and C. Heath, Telling Half the Story: A Critical Review of Research on the Teaching Beliefs and Practices of University Academics, *Review of Educational Research*, **72**(2), pp. 177–228, 2002.
42. N. L. Leech and A. J. Onwuegbuzie, An array of qualitative data analysis tools: A call for data analysis triangulation, *School Psychology Quarterly*, **22**(4), pp. 557–584, 2007.

Paul Bazelais is a doctoral student at McGill University and an instructor at John Abbott College in Montreal, QC.

David John Lemay PhD, is a research associate with the Centre for Medical Education, McGill University, QC.

Tenzin Doleck is a postdoctoral research scholar at the University of Southern California.