RESEARCH ARTICLES

"Testwiseness" Among International Pharmacy Graduates and Canadian Senior Pharmacy Students

Anisah Mahamed, BSc, Paul A.M. Gregory, MLS, and Zubin Austin, PhD

Leslie Dan Faculty of Pharmacy, University of Toronto.

Objective. To compare the test-taking skills and abilities (*testwiseness*) of Canadian senior-level pharmacy students with those of international pharmacy graduates.

Methods. A 20-item testwiseness questionnaire was developed and administered to 102 participants: 35 senior-level pharmacy students, 34 international pharmacy graduates, and 34 practicing pharmacists who served as a control group.

Results. Mean testwiseness scores indicated significant differences in performance between senior-level pharmacy students and international pharmacy graduates. Testwiseness deficiencies of international pharmacy graduates were particularly severe in domains requiring discerning use of English language.

Conclusions. Differences in testwiseness appear to exist between Canadian senior-level pharmacy students and international pharmacy graduates. The genesis and implications of these differences must be evaluated further in order to determine whether testwiseness affects learning, professional development, or clinical practice.

Keywords: examination, testing, test-taking strategies, assessment, international pharmacy graduates

INTRODUCTION

Well-constructed multiple-choice tests can be an effective, efficient, reliable, and valid mechanism for assessing knowledge and skills.[1-3] As such, they have long been favored as a means of assessing individuals in a variety of domains, ranging from academic settings to professional certification processes, as well as in everyday life, such as obtaining a driver's license.[4-6] Despite their popularity, these tests provoke much debate over whether they can truly examine application of knowledge, particularly in a clinical field such as pharmacy[7] or medicine.[8]

Testwiseness has been defined by Gibb as "...the ability of a (test taker) to react to the presence of secondary cues in ways advantageous to himself on a multiple-choice test of knowledge of factual information."[9] Since first being described by Gibb in the early 1960s, testwiseness has been a source of considerable concern to teachers – and amusement to students. The notion that it may be possible for a student to outwit a standardized test and perform well despite a significant lack of content-

specific knowledge runs counter to principles of effective assessment.[10,11]

Students who are *testwise* are able to look for errors in the construction of test items, particularly in multiple-choice questions.[11] Students who are able to outwit a test receive scores that are not valid, and not predictive of their current knowledge and skills or future abilities.[4,10] It is important to differentiate between testwiseness and educated guessing. Testwiseness is based on little or no content knowledge and is merely an attempt to select the correct answer based on errors in test construction. In contrast, making educated guesses, requires the student to have some measure of content knowledge, enough at least to rule out some plausible distractors, reducing the number of possible answers from which a guess may be made.[9,11]

Since first being identified as a threat to the validity of multiple-choice tests, numerous guidelines have been published providing teachers with important "tips" for designing multiple-choice assessments to circumvent common testwiseness strategies (Table 1). Proponents of this approach support the notion that effective test design can prevent successful application of testwiseness.[12] Such design is an important part of the development of high-stakes examinations, such as those governing entry to practice in health professions.

Corresponding Author: Prof. Zubin Austin. Address: Leslie Dan Faculty of Pharmacy, 19 Russell Street, Toronto ON, M5S 1A1. Tel: 416-978-0186 E-mail: zubin.austin@utoronto.ca

Table 1. Testwiseness Strategies Used by Students Taking Multiple-Choice Examinations

Strategy	Description	Example
Order of answer	Recognition of predictable patterns for answer options, usually placing the correct answer in a defined position	Avoiding "A" or "D" as the correct answer and preferring "B" or "C"
Grammatically correct stem	Grammatical accord between step and answer options (eg, correct article, gender accord, plural, grammatical tense, etc)	Use of "an" at the end of a stem, followed by answer options that begin with a consonant will be eliminated by the student
The longest answer option	In an effort to produce an undeniably correct answer, the length of the correct option may be out of proportion to other options provided	To ensure an answer is unambiguously correct, it may be necessary to include so much information the length of the statement grows out of proportion.
Strong modifiers	Use of definitive, highly certain terms require a student to only produce one counter-example to refute the option	Use of words such as "All" or "Always" or "Every" as opposed to "Some," or "Most," or "Often"
Excess specificity	Very detailed options that provide no room for ambiguity may appear to be a correct answer.	Inclusion of highly specific dates, times, or facts
Implausible distractors	Students may be able to select the correct answer by discarding implausible distractors rather than actually knowing the correct answer	Writing 3-4 plausible distractors is frequently difficult. Consequently, some distractors may be less "believable" and thus easier to eliminate without any content knowledge

Although teachers now have resources available to prevent students from succeeding academically through testwiseness alone, for a variety of reasons, testwiseness may still allow some students to over-perform on some tests relative to their actual abilities."[13] In large part this is due to the logistical difficulties associated with developing testwise-proof tests; large, multi-centre high-stakes examinations can afford to invest in developing test items that follow recommendations to improve validity. However, individual teachers who are juggling multiple priorities may simply not have the time or expertise to develop testwise-proof assessments.

As stated by Gibb, the challenges associated with constructing valid multiple-choice test items are considerable.[9] Though no systematic study has been reported to determine the extent and prevalence of testwise-proof tests in postsecondary education, anecdotal reports from students suggest that testwiseness is still an important skill and effective strategy in test taking.[13]

Testwiseness may be a learned behavior that is reinforced and improved over time and with repeated exposure to multiple-choice tests at the high school and university levels.[4] Simply put, the more poorly constructed multiple-choice tests one takes, the more attuned one becomes to the patterns that appear to underlie successful testwiseness. Testwise advice like "when in doubt, pick C", and "if you don't know, pick the longest answer" is still passed down from generation to generation of students.

There is no reason to believe that testwiseness is particularly helpful or effective in high-stakes examination settings such as national licensing examinations in health professions since these examinations have resources to ensure items are testwise-proof. However, in many other academic settings and in other walks of life, multiple-choice tests may be an important factor in advancement or promotion.[6] Within pharmacy education, there has been no work reported on the prevalence or impact of testwiseness. Since pharmacy students were generally among the most academically successful students in postsecondary education, it is reasonable to assume they have acquired some degree of testwiseness during their education. One particular cohort of pharmacy students may not, however, have the same level of testwiseness. For the purpose of this article, international pharmacy graduates (IPGs) are internationally educated health care professionals seeking licensure in Canada.[14] These individuals have completed their academic preparation,

in-service training, and licensing procedures in a country other than Canada or the United States. As part of the requalification process in Ontario, Canada, these individuals are required to complete bridging education courses offered at the postsecondary level. As part of the assessment of these courses, IPGs may be required to complete course-specific multiple-choice tests, where testwiseness skills might prove to be advantageeous. Consequently, studying the testwiseness skills of these students is of interest in understanding the validity of examination results.

The purpose of this study was to compare testwiseness skills of IPGs, and senior-level Canadian pharmacy students (ie, those in professional years 3 and 4) to determine how testwiseness skills compare between these 2 cohorts. To control for potential age-related differences in performance, a cohort of practitioners of similar age/experience to the international pharmacy graduate group was also included in the study.

METHODS

A test blueprint was created, based on the cueing strategies described in Table 1. Questions were created for each cueing strategy. In order to ensure this study did not inadvertently examine domain-specific content knowledge, test items were created that were deliberately content free; this was not designed in any way to be a test of pharmacy-specific knowledge or skills. To ensure consistency, all test items were constructed using a standard stem formation followed by 4 distractors. Appendix 1 provides examples of test items and the cueing strategies they are meant to depict. In order to assess testwiseness skills, Gibb[9] suggests avoiding the use of real content, since this may interfere with the ability to actually measure testwiseness. Instead, construction of "nonsense" items with fictitious words or phrases grouped in familiar patterns should be used to specifically evaluate testwiseness. Based on this approach, a "correct" answer is one that is based solely on correct application of a testwiseness principle, as illustrated in Appendix 1. Using this approach, there is no risk that content-knowledge will interfere with measurement of testwiseness skills, since there is no real content being tested.

A total of 31 questions were developed for all cueing strategies. A validation process was utilized using 5 volunteer students and IPGs. Based on this process, 11 questions were discarded due to problems with readability or non-applicability to the cueing strategy. As a result, a 20-item test was developed. A test of this length balances the ability to utilize multiple testwiseness strategies on several occasions within a reasonable time period.[4,5,9] Previous tests of testwiseness have ranged considerably in length, but most recent examples have typically contained 20-25 items.[11]

A sample size for the number of participants was calculated based on a confidence of 95% ($\alpha = 0.05$) and a power of 80% ($\beta = 0.2$); a minimum of 34 participants from each cohort was required to conduct this study.[15]

Participants for this study were recruited from the senior-level pharmacy students and a cohort of IPG students enrolled in a bridging education program at the University of Toronto. Participants were advised of the purpose of the study and invited to complete the test following completion of signed consent. During the information session provided to all potential participants, questions related to the use of data, anonymity of participants, and rationale for the study were discussed in depth to ensure that all participants were fully aware of their role in this study. In addition, investigators provided opportunities for potential participants to contact them prior to commencement of the study.

Following administration of the test to the IPG group, demographic data were analyzed and a matched group of practicing pharmacists with similar age and experience characteristics were recruited to participate.

Test scores were analyzed grouping the sample into 3 cohorts: senior-level students, IPGs, and pharmacists. Mean test scores with standard deviations were calculated by assigning a value of 1 for a correct response and 0 for an incorrect response, divided by the total number of test items (20), then converted to a percentage. One-way analysis of variance (ANOVA) was used for multiple comparisons ($p = 0.05$), based on mean scores attained by each of the cohorts. For comparisons with only 2 groups, independent samples *t* tests were used. To more clearly compare and contrast differences between the IPG cohort and the senior-level student group, multiple and separate 2-way comparisons were performed and reported (rather than a more traditional 3-way ANOVA followed by post hoc analysis such as Tukey's HSD or Scheffe's procedure).[15] *SPSS* v.11.0 for Windows was utilized for data analysis. Ethics approval for this study was sought and received through the University of Toronto's Ethics Review Board.

RESULTS

Thirty-five senior-level pharmacy students, 34 IPG students, and 34 pharmacists participated in this study. All participants were volunteers; senior-level pharmacy students and IPG students were invited to participate through active recruitment in a captive classroom situation. Pharmacists involved in this study were recruited at continuing education events and invited to participate after completion of signed, informed consent. All participants completed all 20 items of the test. Twenty-eight

Table 2. Results of a Multiple-Choice Test to Identify Testwiseness Among International Pharmacy Graduates and Senior Pharmacy Students

Testwiseness Strategy	Score, Mean (SD)		
	Senior Students (n =35)	IPGs (n = 34)	Pharmacists (n = 34)
Overall mean test score, 20 questions	16.4 (2.6)	10.8 (3.8)	15.1 (1.5)
Order of answer (3 questions)	2.2 (0.5)	1.6 (0.8)	2.0 (0.2)
Grammatically correct stem (4 questions)	3.3 (0.4)	2.1 (0.6)	3.1 (0.3)
Longest answer option (3 questions)	2.1 (0.4)	1.4 (0.7)	1.9 (0.3)
Strong modifiers (3 questions)	2.6 (0.3)	1.6 (0.6)	2.4 (0.4)
Excess specificity (4 questions)	3.3 (0.4)	2.1 (0.7)	2.9 (0.3)
Implausible distractors (3 questions)	1.9 (0.6)	1.5 (0.8)	1.7 (0.5)

IPGs = international pharmacy graduates

percent of senior-level students were male, compared with 59% of IPG students, and 54% of pharmacists. The mean age of senior-level students completing the test was 24 years, compared with 39 years for the IPG students and 40 years for pharmacists.

Tables 2 and 3 present results and cohort-specific comparisons. The only significant difference in mean performances was between the senior pharmacy student cohort and the IPG cohort (82% vs 54%); while the performance of practicing pharmacists was slightly lower compared with that of senior pharmacy students, and slightly higher compared with that of IPGs, this difference did not reach significance

Table 3. Differences in Testwiseness Among Senior Pharmacy Students International Pharmacy Graduates, and Pharmacists

Testwiseness Strategy	Comparison Across All Groups*	Comparison Between Senior Pharmacy Students and International Pharmacy Graduates[†]	
	P	t(69)	P
Order of answer	0.24	1.81	0.23
Grammatically correct stem	0.46	0.99	0.30
Longest answer option	0.26	1.74	0.25
Strong modifiers	0.38	0.94	0.29
Excess specificity	0.42	0.92	0.37
Implausible distractors	0.19	1.09	0.79

*one-way ANOVA
[†]independent samples *t* test

Data were analyzed based on cueing strategies utilized. Once again, the only significant differences between cohorts were between the senior pharmacy student cohort and the IPG cohort for the following 3 cueing strategies: grammatically correct stem, strong modifiers, and excess specificity. There were no significant performance differences between the student and pharmacist cohorts, or the IPG and pharmacist cohorts.

DISCUSSION

This study suggests that testwiseness is a well-developed skill among pharmacy students who have had their primary and secondary education in North America. Across all types of testwiseness skills assessed, the majority of these students were able to discern the ''correct'' response, presumably by recognizing and responding to specific cueing strategies built-in to each question.

International pharmacy graduates, however, were less successful in recognizing and responding to these cueing strategies, and in particular strategies requiring them to utilize sophisticated or subtle English language fluency skills, including grammatically correct stem, excess specificity, and strong modifiers. Given the design of this study, it was not possible to determine whether performance differences between the IPG group and the senior-student group are a reflection of underdeveloped English language skills, lack of experience with multiple-choice testing, or lower testwiseness skills, since differences in performance between the IPG cohort and the control group of pharmacists did not reach significance.

An important consideration in this study is previous experience with multiple-choice testing formats. As

Sarnacki[4] has noted, previous exposure to poorly constructed tests may "prime" students' testwiseness skills and make them more conscious of cues and more successful in outwitting tests in the future. This study did not attempt to control for educational background of participants in the IPG cohort, some of whom may not have had any experience with multiple-choice testing methods in their pharmacy education. Such differences in academic preparation may have affected testwiseness results observed in this study.

Attempts to study testwiseness suffer from several important limitations. First, the inauthentic conditions of this study directly effect a student's motivation. Consequently one cannot truly judge performance based on this sort of simulation. All participants in this study had no particular reason to care about its outcome, since the results will not affect them personally. While in a real testing situation, a student would have ample incentive (in the form of course grades) to apply testwiseness or other skills to finding a correct answer, there is neither incentive nor reason to do so under the conditions of this study. This type of study (using nonsense questions) has the advantage of neutralizing any content or knowledge advantage; however, it may also diminish incentives to perform effectively on the test.[4] This limitation notwithstanding, one can reasonably conclude this affected all cohorts in the study equally and consequently between-group comparisons should not have been affected significantly.

Second, there is no way of determining whether testwiseness is an issue for the IPG cohort, or instead, if English-language proficiency or cultural competency is the more important reason underlying performance differences. All IPG participants in this study met minimal English-language fluency requirements for practice as a pharmacist in Canada. Undoubtedly, the level of English language proficiency of the IPG cohort was qualitatively lower than that of the senior-pharmacy student or practicing-pharmacist cohorts. Nevertheless, all IPG students passed standardized, objective English tests designed to ensure they were capable of meeting English language demands of pharmacy practice. This, however, raises the intriguing question of whether "minimal" language requirements for pharmacy practice are not discerning enough to allow testwiseness issues to emerge.

Third, the design of the instrument used to measure testwiseness may have also introduced limitations. As discussed previously, the 20-item questionnaire used in this study was meant to balance competing needs to ensure adequate data capture without being so lengthy as to alienate or bore potential participants. At 20 items, it was only possible to use 2-4 different questions for each testwiseness strategy. With such a small sample for each strategy, it may be difficult to draw strong conclusions regarding participants' performance. While a larger number of items enhance statistical analysis, it may paradoxically affect performance by increasing boredom or disengagement from the task, and was thus rejected for this study.

This study does not address a particularly salient question: how does testwiseness emerge? While there is some conjecture that it is a learned, behavioral response to poorly constructed multiple-choice tests there is no evidence to support this, particularly in the context of pharmacy education. However, this study confirms that testwiseness exists among senior-level pharmacy students educated in North America, and appears to be less prevalent among international pharmacy graduate students. The implications of this finding for educators, researchers, regulators, and students themselves need to be fully evaluated in the context of emerging trends in pedagogy.

CONCLUSIONS

This study suggests that testwiseness skills are prevalent among North American students and less prevalent among international pharmacy graduates. Further work is required to elucidate mechanisms for development of testwiseness skills in different groups. In particular, additional research is needed to determine whether high testwiseness scores correlate with performance in experiential learning, in clinical practice, in retention and recall of learned material, or in day-to-day practice as a pharmacist.

REFERENCES

1. Beullens J, Van Damme B, Jaspaert H, Janssen PJ. Are extended-matching multiple-choice items appropriate for a final test in medical education? *Med Teach.* 2002;24:390-5.
2. Diamond JJ, Evans WJ. An investigation of the cognitive correlates of testwiseness. *J Educ Meas.* 1972;9:145-50.
3. Damjanov I, Fenderson BA, Veloski JJ, Rubin E. Testing of medical students with open-ended uncued questions. *Hum Pathology.* 1995;26:362-5.
4. Sarnacki RE. An examination of testwiseness in the cognitive test domain. *Rev Educ Res.* 1979;49(2):252-79.
5. Slatker MJ, Koehler RA, Hampton SH. Learning testwiseness by programmed texts. *J Educ Meas.* 1970;7:247-54.
6. Millman J, Bishop CH, Ebel R. An analysis of testwiseness. *Educ Psych Meas.* 1965;25(3):707-25.
7. Stupans I. Multiple choice questions: can they examine application of knowledge? *Pharm Educ.* 2006;6:59-63.

8. Schuwirth LW and van der Vleuten CP. ABC of learning and teaching in medicine: written assessment. *BMJ.* 2003;326:643-5.

9. Gibb BG. Testwiseness as secondary cue response. (Doctoral dissertation, Stanford University) Ann Arbor, Mich: University Microfilms, No. 64–7643.

10. Miller PM, Fagley NS, Lane DS. Stability of the Gibb (1964) experimental test of testwiseness. *Educ Psych Meas.* 1988;48: 1123-7.

11. Geiger MA. An examination of the relationship between answer changing, testwiseness, and examination performance. *J Exp Educ.* 1997;66:49-58.

12. Kehoe J. Basic item analysis for multiple choice tests. *Practical Assess Res Eval*. 1995;10(4). Retreived January 11 2006 from: http://PAREonline.net/getvn.asp?v=4&n=10.

13. Fenderson BA, Damjanov I, Robeson MR, Veloski JJ, Rubin E. The virtues of extended matching and uncued tests as alternatives to multiple choice questions. *Hum Pathol.* 1997;28:526-32.

14. Austin Z, Galli M, Diamantouros A. Development of a prior learning assessment for pharmacists seeking licensure in Canada. *Pharm Educ.* 2003;3:87-96.

15. Moore DS, McCabe GP. *Introduction to the Practice of Statistics,* 4th ed. New York: W.H. Freeman and Company, 2003.

Appendix 1. Sample Testwiseness Instrument Questions

1. Grammatically Correct Stem

The use of maxamolol as a replacement therapy may treat an:

a) progesterone excess
b) androgen deficiency
c) corticosteroid excess
d) leukotriene deficiency

Correct answer is "b". Note use of "an" in stem, which grammatically is correct only with "androgen."

2. Longest Answer Option

Which of the following statements about Quikofelbads is incorrect?

a) Effective treatment is bedrest and drinking plenty of fluids.
b) Zingonine is an antiviral treatment that may cause nausea.
c) It can undergo mutations.
d) Cases of respiratory syncitial virus (RSV) are prevalent among the elderly but are seldom reported to public health services.

Correct answer is "d". Note very detailed response that requires a disproportionately long answer.

3. Use of Strong Modifiers

Hapincantin:

a) should be taken with food to prevent nausea
b) must be taken at least two hours after any cardiac medication
c) cannot be taken with Chanto-Berchunin
d) will reduce effectiveness of birth-control pills

Correct answer is "a". Note use of definitive modifiers such as "must", "cannot" or "will" in other distractors vs. "should" in "a" which appears more measured.

4. Excess Specificity

In 2001, researchers isolated and sequenced euph88, a potential candidate for gene therapy to treat adolescents with mood disorders. The DNA of euph88:

a) is photosensitive using normal photosensitivity analysis methods, unlike its analogue euph 82
b) has a wavelength of 460 nm, as measured using normal photosensitivity analysis methods
c) has a sequence homology with Nop genes which are associated with neurodegenerative symptoms in pregnant mice
d) is degraded using a mild detergent

Correct answer is "c". Note seemingly redundant and non-specific information in distractors "a" and "b" and non-specific distractor "d".