Use of a Diagnostic Tool to Predict Performance on High-Stakes Multiple Choice Tests: A Case Study

Natalie N. Michaels, Timothy Stewart, Ronald Barredo,
Edilberto Raynes, Deborah Edmondson, and Elizabeth Kunnu

Abstract

High-stakes testing can be a major hurdle for individuals who know the material well, but have trouble understanding the language of the test. Many people have difficulty understanding test questions when the wording of the question is different from the language variation typically used by the test-taker. This research builds on prior research published in 2011 and 2016. The Language Variation Tool (LVT) was created to help determine if problems existed when test questions were reworded in accordance with Southern Caucasian and Southern African American language variations. This tool focused on physical therapy graduate students, and a study published in 2016 demonstrated some predictive value to the LVT with respect to the PT board practice examinations. Students who received a lower overall score on the diagnostic test, also demonstrated lower scores on the practice board examination (Pearson correlation=.662; p=.000). This current paper focuses on how the tool was created, and looks at one of many students who have used this tool to help pass the physical therapy board examination. When there is a misunderstanding of the meaning of test questions, there can be increased difficulty passing standardized examinations. This case study supports the utilization of this diagnostic tool to predict success on the high stakes examination, and the power of intervention to overcome the obstacle of successful comprehension. These authors believe that this tool can be created in other languages for other disciplines focusing on various cultural proclivities. Because this is a simple case study, the results may not be generalizable to the population of those taking the physical therapy board examination. As such, it is the opinion of the researchers that simply informing a test-taker of his or her limitations prior to an exam, gives them something to practice and develop prior to retaking the examination with the hope of potentially improving test performance. This contention would require further research.

About the Authors: Natalie N. Michaels, PT, EdD is a Professor in the Physical Therapy and Occupational Therapy Departments at Belmont University, College of Health Sciences, and a Geriatric Clinical Specialist Emeritus through the American Board of Physical Therapy Specialties. Timothy Stewart, DPT is a Physical Therapist with the Tennessee Orthopedic Alliance in Hendersonville, TN. Ronald Barredo, PT, EdD, DPT,GCS is a Professor and Chair of the Department of Physical Therapy at Tennessee State University (TSU), Edilberto Raynes, MD, PhD is an Associate Professor in the Physical Therapy Program at TSU, Deborah Edmondson, PT, EdD is a Professor in the Physical Therapy Program at TSU, Elizabeth Kunnu, Ed.S., M.Ed., RHIA, was an Assistant Professor and Chair of the Department of Health Information Management and Respiratory Care at TSU at the time of this study.

Introduction

High-stakes testing can be frustrating for students who are confident in their knowledge of the test content but are still unable to pass the examination. One potential reason for this difficulty is a misunderstanding of the test questions. This research utilizes the Language Variation Tool for physical therapist students/graduates (LVT-PT), and builds on work that was previously published in the Journal of Interdisciplinary Education in 2011 and 2016 (Housel et al., 2011; Michaels et al., 2016).

The Problem

Language and communication skills have been cited as major contributors to student success and attrition (Bosher & Smalkoski, 2002; Bosher, 2003). Performance on a high-stakes examination that is verbose may be more dependent upon reading comprehension than content knowledge. (Haladyna, 2004; Schellenberg, 2004). Students who demonstrate advanced knowledge in course content, will still sometimes report difficulty passing multiple choice examinations, simply because they misunderstand the test questions. Although a multitude of interacting variables could lead to discrepancies in test scores, this research focuses on English language variation of multiple choice questions.

The 2011 Study

In 2007, Housel and associates developed a diagnostic multiple-choice testing instrument called the language variation tool (LVT-PT), to help determine whether or not a student in a graduate school physical therapy program was affected by the wording of questions during multiple choice testing. A pilot study was performed using the LVT-PT at Tennessee State University (TSU) for the doctoral physical therapist (DPT) graduating classes of 2008, 2009 and 2010 (Housel, 2011). Although it was found that there were some students from all cultural proclivities who had difficulty answering questions when they were re-worded, there was a higher percentage of African American students with this problem (with an intercept of p=.000).

The 2016 Study

It became important to know the predictive effectiveness of the LVT-PT for future high stakes testing, specifically the National Physical Therapy Examination (NPTE). A study conducted by Michaels et al. (2016) demonstrated that there was predictive value to the LVT-PT. The creators of the NPTE have published a practice examination called the Practice Examination Assessment Tool (PEAT) that uses language very close to that of the NPTE (FSBPT, 2015). The PEAT was found to correlate highly with the NPTE results by Barredo, Tan, and Raynes (2015, preliminary result). Because the scores from the PEAT were more convenient to access than the scores for the NPTE, the PEAT was used in this study. There was a moderately strong correlation when comparing the scores of the LVT-PT to the PEAT. (Pearson correlation=.662; *p*=.000). This correlation was stronger than that of other practice tools that don't use the specific language variation used in the PEAT and the NPTE (correlation=.392; *p*=.029). Problems with language variation during multiple choice testing may be more common than originally thought. During the earlier study, 22 of the 35 students scored a minimum difference of 15% or higher on the LVT (Michaels et al., 2016). This was 62.8% of the class (Figure 1).
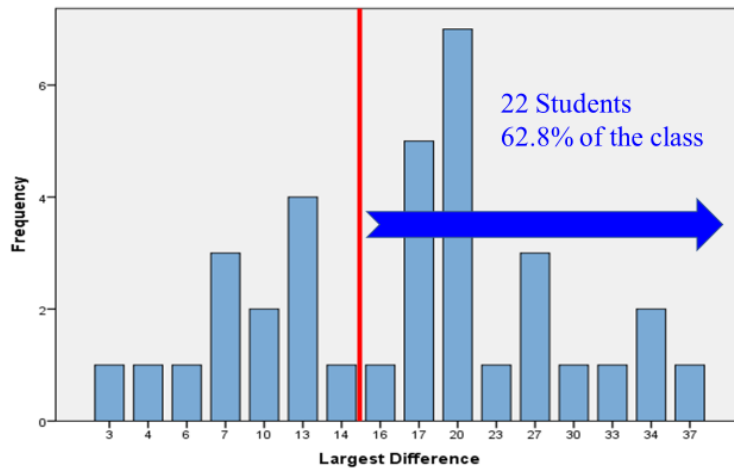
22 Students
62.8% of the class

***Figure 1*. The largest difference in scores when questions were re-worded (Michaels, et al. 2016; used with permission).**

Importance of Intervention

The first step to intervention is feedback. Butler and Roediger (2008) found that the provision of feedback after testing led to an increased proportion of correct responses on later testing. Once success can be predicted, feedback can be provided, and then assistance can be implemented.

Method

The LVT was created in 2008, with an outline created from Domholdt's *Rehabilitation Research: Principles and Applications, 3rd ed.* (Domholdt, 2004). The recommended steps include drafting, expert review, a first revision, a pilot test, and a final revision. Although the creation of the LVT is quickly summarized in this article, a more comprehensive depiction of tool's development is outlined in the 2011 article (Housel et al., 2011). The authors believe that the same strategy could be followed to potentially re-create this tool for other disciplines in various languages as a to way help detect students who have problems with language variation during multiple choice testing.

Drafting

Test questions were obtained from a preparatory guide for the NPTE with permission from the author, Scott Giles (Giles, 2007). A group of 6 students (3 Caucasian, and 3 African American), volunteered to help re-write questions "as though they were talking to their friends." Thirty of the re-worded questions (15 from the Caucasian students and 15 from the African American students) were then mixed into a test with 30 more questions as originally worded by Giles. A second test was also created, re-wording the opposite 30 questions, to eliminate the variable of question difficulty when looking at group differences.

Expert Review

The re-written questions were first reviewed by the faculty research team, then sent to two expert linguists for review. For the Caucasian-worded questions, the expert was Walt Wolfram, PhD, is a William C. Friday Distinguished Professor of English Linguistics, North Carolina State University, Director of the North Carolina Language and Life Project, co-creator of the Interinstitutional Cooperative PhD English Linguistics Program at Duke University and author of *American English: Dialects and Variation* (Wolfram,1998). For the African American-worded questions, the expert was Lisa Green, PhD, is a

Professor at the University of Massachusetts Amherst, and author of *African American English, A Linguistic Introduction* (Green, 2002). These experts read and revised the questions to ensure that the items actually reflected the language variation in question.

First Revision

After the expert review, the items were returned to the faculty to check that the meaning of the questions remained unchanged. The questions were also checked by the faculty for content validity by answering the following questions: (1) Were the major elements addressed? (2) Were the questions understandable within the limits of the dialect? (3) Were the terms defined satisfactorily? (4) Were the questions formatted appropriately? As per Domholdt (2004). Questions were revised accordingly.

Pilot Test and Final Revision

Upon TSU Institutional Review Board (IRB), the pilot study was conducted in September, 2008, with the TSU physical therapy graduating class of 2008 (*N=17*), the class of 2009 (*N=19*), and the class of 2010 (*N=22*). The LVT was given to each student on two separate days; half taking Test A on day one, and half took Test B, then switching on day two. Differences were found in many student test scores for both Caucasian and African American students when questions were reworded, but the differences were more evident in the African American student group when compared to the Caucasian group (with an intercept of p=.000), regardless of whether test A or test B was used (Housel et al., 2011). The Caucasian students scored higher for everything, except the African American worded questions (Table 1), which were scored highest by the African American students.

	Caucasian	African American
Total Score %	61.59	59.39
Caucasian % Score	59.45	56.19
African American % Score	57.23	64.27
Originally Written % Score	62.71	58.59

Table 1: **Group data showing average results of two days of testing (From Michaels et al., 2016; Used with permission)**

Grade Interpretation

The cut-off for potential problems passing the board examination appeared to be when students demonstrated a 15% or higher score difference between questions as originally worded and questions that were re-worded.

The Case

The following depicts an actual case study depicting one of the many success stories resulting from the LVT-PT, a Caucasian male in his late 20s. We will call him Joe. Joe graduated with his doctorate in physical therapy 1.5 years prior to receiving this intervention. He had taken the NPTE 5 times without success. Passing score at the time was a grade of 600, and he consistently received a score of around 595. He came into the college and met with Dr. Natalie Michaels who gave him the LVT-PT. Joe demonstrated a 23% score difference between questions as originally worded and questions that were re-worded. This was 8% higher than the cut off for possible problems with language variation.

Intervention

Joe was immediately provided feedback of his test results, and the potential problem was explained to him with some ideas for intervention to be done prior to re-taking the NPTE. It was decided that Joe would take the PEAT, then meet with Dr. Michaels four to six times to go over the meaning of questions missed. This would continue until Joe was able to pass the PEAT two times in a row. During the meetings with Dr Michaels, missed questions on a computer screen would be read out-loud. For each question, Dr. Michaels would ask Joe to explain in his own words what the question was asking. Both would then choose and answer each question based upon their perceived interpretation of the question. It was obvious that Joe had a good command of the content, but that he was unclear of the meaning of the questions, however, Joe's performance improved with each subsequent meeting until he could easily pass the PEAT. He also demonstrated more confidence with each subsequent meeting.

This took six months, and five meetings with Dr. Michaels, during which time, Joe performed clinical observation hours "to stay fresh." He also met with a couple other professors to make sure he still knew the content.

Results

When Joe took the exam for the 6th time, he received a passing grade of 649. The following questions were later answered later by Joe:

1. Tell me what you went through: "*I was frustrated more than anything. I was depressed for a while. I thought that I had spent this many years to get to this point, and now, I'm basically useless.*"

2. What was that process (with Dr. Michaels) like for you? *"The language variation sessions more than anything made me start to wonder, what does the question really ask?"*

3. How did it feel knowing the reason you had not yet passed? "*For me, it felt like I wasn't crazy. I did know the content. If I can pass a test, when it's worded differently, then it's obvious it's not a content issue. I took the practice classes in Orlando, and I knew the content. You are doubting yourself, but once you figure out what they are asking, and you pass a practice exam, you realize, you know what the hell you're doing now, so now, go put it to use!*"

4. How did you pass the last time,? What did you change? *"I focused on the language, and that was it. I didn't really study the content the last time, and I also didn't tell anyone I was taking the exam. I kept that to myself, my wife and the professors at TSU. I didn't tell anyone else. I also allowed myself to take more frequent rest breaks."*

5. Do you feel that this system is worth revisiting to help benefit others? *"I'm working at a good salary at good company, so yes, this was really worth it!"*

Discussion and Conclusion

Many students who perform well in their college courses, will have trouble passing high-stakes board examinations. There can be many variables causing this phenomenon, but this paper supports Bosher & Smalkoski (2002) contention that good language and communication skills are paramount. This study also supports the work of both Haladyna, (2004) and Schellenberg (2004), who believe that a test can sometimes be more a measurement of reading comprehension than knowledge of the subject matter. Understanding

the question is the first step to answering it correctly. Some students have difficulty with this, and this research has found one possible way to detect this in the student population.

Because of the many variables that can affect performance on examinations, it's difficult to tell how many may have affected Joe's performance. Even his increase in confidence might have been enough to push him over the threshold, but clearly, knowing the language of the test could have made a difference in his level of confidence. Whatever the case, Joe went into the 6th examination realizing that one of his limitations was controllable and knowing what to do to eliminate the obstacle. He approached the examination armed with a new effective strategy for success.

Limitations
Regardless of this success story, there are limitations. The LVT was only developed for students and graduates in a doctoral physical therapist program prior to the board examination. This tool needs to be developed and tested to determine its effectiveness with other disciplines and cultures. The authors therefore warn against generalizations to any other population.

Clinical Impact
Problems with language variation during multiple choice testing may be more common than originally thought, as found in the study by Michaels et al. (2016). Coupled with the finding that it appears more common in minority populations is also of interest. Knowing who has the problem and providing intervention could prove beneficial, not just for physical therapy but for any discipline that requires a high-stakes board examination. According to the Federation of State Boards of Physical therapy, 86% of first-time test takers passed the board examination in 2018 (FSBPT, 2018). How many of the 14% who did not pass simply had difficulty understanding the meaning of the questions? Another example could be nursing. According to the National Council on State Boards of Nursing, 89.94% of first-time test takers passed the exam in 2019 (NCSBN, 2019). The reason for those who did not pass is unknown. Some would assume that this is due to a lack of required knowledge and/or inability to apply such knowledge. Some of the test-takers, however, might have simply have had a problem understanding the questions. After all, the students acquired enough knowledge to complete the didactic and clinical portions of the curriculum. Clearly, some other issues are interfering with their success on the board examination. Future investigation is necessary.

Conclusion
Although the simple case study is not entirely generalizable to the population of those taking the physical therapy board examination, it is one case of many. This research has focused on one possible limitation to success when taking high-stakes examinations, inability to understand some or all of the test questions, due to language variation. It is the opinion of the researchers that informing a test-taker of his or her limitations prior to an exam, gives them something to practice and develop prior to retaking the examination, and could potentially improve test performance. This contention would require further research.

References

Barredo, R., Tan, J., & Raynes, E. (2015). GPA, PEAT scores, and the NPTE: A pilot Study. Abstract Submission 2016 CSM. *American Physical Therapy Association.* Available through Ron Barredo, Tennessee State University.

Bosher, S. (2003). Barriers to creating a more culturally diverse nursing profession: Linguistic bias in multiple-choice nursing exams. *Nursing Education Perspectives, 24*(1) 25-34.

Bosher, S., Smalkoski, K. (2002). From needs analysis to curriculum development: Designing a course in health-care for immigrant students in the USA. *English for Specific Purposes, 21*(1)59-79.

Butler, A.C., Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*(3), 604-616.

Domholdt, E. (2004). *Rehabilitation Research: Principles and Applications, 3rd ed.* Saint Louis, Missouri: Elsevier Saunders.

FSBPT (2015). The Practice exam & assessment Tool (PEAT). *Federation of State Boards of Physical Therapy* Retrieved: https://www.fsbpt.org/OurServices/CandidateServices/PracticeExamAssessmentTool(PEAT).aspx

*FSBP*T (2018). Pass Rate Reports. *Federation of State Boards of Physical Therapy.* Retrieved. https://www.fsbpt.org/Free-Resources/NPTE-Pass-Rate-Reports

Giles, S. M. (2007). *PT EXAM: The Complete Study Guide*. Scarborough, Maine: Scorebuilders.

Green, L. J. (2002). *African American English: A Linguistic Introduction*. Cambridge, United Kingdom: Cambridge University Press.

Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items (3rd. Ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Housel, N., Barredo, R., Edmondson, E., Raynes, E., Kunnu, E. (2011). Social Justice for multiple choice examinations: Development of a diagnostic tool to detect student problems with language variation. *The Journal of Interdisciplinary Education*, *10*(1), 92-107.

Michaels, N., Barrredo, R., Raynes, E., Edmondson, D., Kunnu, E. (2016). A Diagnostic Tool to Predict Performance on High-Stakes Multiple Choice Tests: An Attempt to Recognize Potential Discrepancies Due to Diversity Before the Test is Taken. *Journal of Interdisciplinary Education,15*(1):51-68.

NCSBN (2019). 2019 NCLEX Pass Rates. *National Council of State Boards of Nursing.* Available. https://www.ncsbn.org/Table_of_Pass_Rates_2019.htm

Schellenberg, S. J. (2004). *Test Bias or Cultural bias: Have We Really Learned Anything?* Paper Presentation for the Annual Meeting of the National Council for Measurement in Education. San Diego, CA. April 14, 2004.

Wolfram, W. (1998). *American English: Dialects and Variation*. San Francisco, CA: Wiley Blackwell.