

Evaluation of Proposals  
for the  
Pharmacy Technician Testing and Certification Services  
to the  
Texas State Board of Pharmacy

Dr. Thomas M. Haladyna  
Professor Emeritus  
Arizona State University

March 31, 2014

This report provides ratings on two proposals and comprehensive discussions of differences in specific ratings for each proposal. The appendix provides the ratings for each proposal with differences highlighted in shade. Also, presented is a discussion of each proposal and a summative evaluation.

### **Total Ratings and Comprehensive Discussion of Differences**

National Health Association–530

Pharmacy Technician Certification Board–554

#### **2f From a psychometric and interpretative perspective, a description of the criteria and procedures in place for documenting the reasons for removal and/or revision of questions if warranted.**

Both proposals give fairly standard discussion of monitoring items. However, a major problem is in the use of items that have great content relevance and performance is very good, the difficulty index (p-value) is very high. As this test produces criterion-referenced interpretations showing the extent of learning, items with high p-values should be retained if the subject-matter experts think the content is important. On pages 47-48, NHA's discussion did not recognize this point. Its discussion was superficial. On pages 40-41, PTCB provides a richer, more detailed discussion that includes the opinions of subject-matter experts. It discusses drift, which is a threat to validity.

Also, using the point-biserial for evaluating distractors can be misleading. For distractors rarely chosen, it is incorrect. Moreover, it is a biased statistic unless the response to distractors is removed before calculating it. Both proposals state that they do this.

#### **2J. The number of test versions or formats proposed for use in each testing setting.**

On page 53, NHA does not make clear how many test forms exist in each testing cycle and how they might be equated. The discussion is too brief and not informative. PTCB provides a more extensive discussion that makes clear its purpose. On page 44, PTCB has a superior plan for test forms. It incorporates field testing of new items in a very effective way. The multiple forms are operational so that item exposure is minimized.

#### **2K. A description of any quality assurance procedures that are in place to ensure that each delivered examination was appropriate from a content and technical perspective prior to the release of scores.**

On page 53, NHA provides an adequate general discussion of quality assurance. On page 44, PTCB provides a more detailed process provides additional information in the appendix. Both appear to have effective quality assurance procedures. However, the attention to detail is

convincing that one company cares more about quality assurance. The worst thing that can happen in any testing program is loss or exposure of test items. Safety and security are very significant features of any high-stakes testing program.

**2L. A procedure in place to control for exposure rate of individual questions to safeguard the security and integrity of the item bank.**

Based on the discussion on page 55 of the proposal, NHA does not understand item exposure. PTCB does. On page 45, although its discussion is brief, it is to the point. They have a system for rotating items in and out of test forms and a large item bank that protects against item exposure. PTCB also has a large item bank, which is another feature that defends against item exposure. NHA's item bank is insufficient for this purpose, and the use of a single form is not very good. Item exposure and test security are related. Item exposure is a serious threat to validity.

**9B. A description of the information that is in the item bank.**

Item history is a very useful feature of an item bank, because items are annually reviewed and often edited or revised. NHA gives a very vague description on page 73. PTCB has a much more sophisticated set of data kept, which is superior to NHA's method.

**9C. A description of the quality assurance procedures that are in place to ensure the accuracy of information in the item bank.**

Again, on page 74, NHA gives a vague answer, and on page 64 PTCB gives a very specific answer backed up by a reference and example in the appendix. If you trust the veracity of the vague answer, both seem to have good quality assurance, but the proposal that provides most specific good information is very convincing.

**9D. A description of how the item banking software interfaces with the test development process.**

On page 75, NHA provides a very short, vague description. On page 65, PTCB names the software and indicates on it is used along with subject-matter experts and other item review functions.

For the most part, the differences in the two proposals seem minor. Both testing organizations appear capable. However, the item bank and how items are developed and validated are crucial to achieving high validity. In this regard, PTCB is the superior organization.

## **Summative Evaluation**

### **NHA**

NHA has prepared an excellent proposal that addresses all the criteria used in this evaluation. The use of a second volume containing appendices is effective in showing that it actually does what it says it does. If NHA was your testing service, it would provide a suitable test development, administration, scoring, and reporting.

However, the item bank is the heart of any testing program. Each item professionally developed has a very sizable market value well more than \$1,000 per item. NHA does not appear to have a large item bank. Moreover, it is not clear how many test forms are developed each year and, if so, how they are equated. Many of their responses were too brief, uninformative, and superficial. One would have to trust that they know what to do and will do it right.

The lower rating seems justified.

### **PTCB**

PTCB has a well-organized, high quality testing program that specializes in this examination. The proposal was well done and detailed in important ways: item development and validation. They have a sizable item bank. They are backed up by vendors/subcontractors with extensive experience and expertise in test administration. This organization exudes confidence that they know what to do. They serve an extensive number of client states.

The higher rating seems justified.

### **Concerns**

Neither organization provides much information about validity. As validity is the most important aspect of any testing program, more attention should be given. Regardless of which testing organization you choose, it benefits you to emphasize with them how their services will improve validity. The technical report is crucial toward the end of analyzing and improving validity. The new *Standards for Educational and Psychological Test* (2014) provide rich detail about validity and test development standards. These too should be emphasized in future work with the testing service you choose.

### **Overall Evaluation**

PTCB is a clear choice. However, both organizations appear highly capable. If, for some reasons, NHA is chosen, it would be very important to iron out details over some of the concerns expressed in this report, especially with respect to test forms, item exposure, equating, and the health of the item bank.

## Appendix

Category	NHA	PTCB
1a	20	20
1b	25	25
1c	25	25
1d	25	15
1e	15	15
1f	10	10
1g	20	20
2a	5	5
2b	5	5
2c	10	10
2d	15	15
2e	15	15
2f	1	5
2g	15	15
2h	15	15
2i	5	5
2j	1	5
2k	8	10
2l	3	15
2m	16	16
2n	10	10
2o	10	10
2p	5	5
2q	10	10
2r	5	5

2s	5	5
2t	5	5
2u	5	5
7a	15	15
7b	15	15
7c	15	15
7d	5	5
8a	6	6
8b	8	8
9a	6	6
9b	9	15
9c	9	15
9d	4	10
10a	10	10
10b	5	5
10c	15	15
10d	15	5
10e	5	5
10f	5	5
10g	5	5
11a	5	5
11b	5	5
11c	3	3
12a	16	20
12b	25	25
12c	15	15
Total	530	554



## Request for Qualifications Summary Report

March 14, 2014

### National Healthcareer Association (NHA)

The proposal submitted by the National Healthcareer Association (NHA) provides a comprehensive response to the scope of work as presented in the request for qualifications (RFQ) authored by the Texas State Board of Pharmacy. The NHA creates and administers certification exams in various fields in the allied healthcare industry, including the Certification of Pharmacy Technicians (ExCPT) exam. Passing of the ExCPT exam leads to NHA's Certified Pharmacy Technician (CPhT) credential/certification. Additionally, NHA's certification is accredited by the National Commission for Certifying Agencies (NCCA) and has been the case since 2008. NHA has issued over 300,000 certifications since 1989. NHA's certifications and training materials are recognized and used by colleges, universities, national chain pharmacies, hospital groups and health care providers across the United States.

The proposal submitted by the NHA includes many positive aspects and several weaknesses. The strengths of the proposal particularly salient to this RFQ include the following:

#### *Strengths*


- Use of a group of subject matter experts (SMEs) that is well-qualified and diverse.
- NHA operates using a continuous quality improvement program approach to test development and delivery.
- NHA uses PSI Services, LLC (a company integrating state-of-the-art test delivery technology and environments) as centers for exam delivery/examinee testing.
- Compliance with the Americans with Disability Act (ADA).
- Test administration and data handling/transmission is secure.
- Develops and administered exams according to guidelines and standards published by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999).
- An organizational infrastructure (e.g., management and technical staff) that provides evidence that NHA will be able to fulfill the obligations outlined in this RFQ if awarded.
- Comprehensive technical manual in accordance to the AERA, APA, NCME Standards for Educational & Psychological Testing (1999) that reports the results of the ExCPT exam post development.
- Strict adherence to standards for certification/licensure testing (and testing practice more generally) as detailed by the American Educational Research Association, American Psychological Association, National Council on Measurement in Education.
- Working a diverse group of subject matter experts and job analysis panelists.
- Clear reporting of the sampling plan (frame), analytic sample used for the job analysis study.
- Item development directed/supervised by appropriately trained employees.
- Adequate item piloting, field-testing, and performance evaluation using classical test theory (CTT).
- Used appropriate techniques for examining differential item functioning (DIF) or "item bias" to ensure that test items are not unfair (psychometrically and culturally) to the various groups being tested.
- Appropriate use of standard setting technique for establishing the cut score for passing the exam.
- Decision consistency analysis of ExCPT to examine the classification accuracy of the exam under actual testing conditions.

- Item response theory based equating of different test forms; this also facilitates interpreting the cut-score across different test forms.
- Thorough and comprehensive item anchoring strategy to facilitate new or parallel test forms using IRT.
- Use of both classical test theory (IRT) and item response theory (IRT) in evaluating item performance (very thorough on this point), test scoring, equating, and evaluating traditional score reliability and IRT-based conditional errors of measurement along the score scale.
- Strong evidence of item bank security (physical, system and security).
- Item bank quality assurance procedures are adequate.
- Strong evidence of high-quality standardization of test administration/delivery.
- Data storage/security procedures are comprehensive and state-of-the art.

#### *Weakness*

- The reliability of rating data (i.e. Intraclass Correlation Coefficient) was not corrected for chance.
- Job analysis not recently conducted.
- The authors state that “descriptive statistics of this size are judged sufficient for the purpose of this job analysis”, however there is no power analysis to verify this claim.
- No sampling error rate (i.e. bias) was reported.
- No factor analysis (or other form of test dimensionality) on the content areas to examine their dimensionality.
- The intraclass correlation coefficients (ICCs) for the item ratings across respondents are high but were not corrected for chance.
- The 1-PL model may not fit these data, and no mention was made regarding IRT model-data fit. So, the *b*-values (item difficulties) and the conditional standard errors of measurement (CSEMs) are likely to be inaccurate. While this may or may not impact the exam results practically, a case must be made for why one particular IRT model over another is to be used (e.g., because the stakes are very high in certification/licensure testing). Choice of IRT model also has implications for test form equating.
- Reporting Coefficient Alpha makes no sense for subtest with fewer than 10 items.
- Regarding stability reliability, the authors state they are in the process of developing a policy for retesting and once instituted stability reliability will become part of the psychometric evaluation of the scores.

Respectfully submitted,

  
Larry R. Price, Ph.D.  
References

Date 3/14/14

## References

- American Educational Research Association/American Psychological Association/National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando, FL: Harcourt, Brace, Jovanovich Publishers.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Belmont, CA: Brooks/Cole Publishers.
- Kline, T. J. B. (2005). *Psychological Testing*. Thousand Oaks, CA: Sage Publications.
- Kolen, M. J., & Brennan, R. L. (1995). *Test Equating: Methods and Practices*. New York, NY: Springer-Verlag Publishers.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory, 4th ed.* New York, NY: McGraw-Hill Publishers.
- Cizek, G. J. (2001). *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Cohen, R. J., & Swerdlick, M. E. (2010). *Psychological Testing and Assessment: An Introduction to Tests and Measurements, 7th ed.* New York, NY: McGraw-Hill Publishers.
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items, 3rd ed.* Mahwah, NJ: Lawrence Erlbaum Publishers.
- Czaja, R., & Blair, J. (2005). *Designing Surveys: A Guide to Decisions and Procedures*. Thousand Oaks, CA: Sage Publications.
- Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E., & Tourangeam, R. (2009). *Survey Methodology, 2nd ed.* New York, NY: Wiley Publishers.
- de Alaya, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York, NY: Guilford Publishers.
- Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. (2007). Standardized Conditional SEM: A Case for Conditional Reliability. *Applied Psychological Measurement*, 31, 3, 169 – 180.

## **Request for Qualifications Summary Report**

**March 14, 2014**

### **Pharmacy Technician Certification Board (PTCB)**

The proposal submitted by the Pharmacy Technician Certification Board (PTCB) provides a comprehensive, state-of-the-art and technically sound response to the scope of work as presented in the request for qualifications (RFQ) authored by the Texas State Board of Pharmacy. The Pharmacy Technician Certification Board is governed by (a) the American Pharmacists Association, (b) the American Society of Health-Systems Pharmacists, (c) the Illinois Council of Health-Systems Pharmacists, (d) Michigan Pharmacists Association, and (e) the National Association of Boards of Pharmacy. The PTCB develops, maintains and administers the Pharmacy Technician Certification Exam (PTCE), an exam that has been administered over 675,000 times. Additionally, 575,000 certifications have been granted and at the end of 2013, over 218,000 individuals possessed an active PTCB credential. The PTCE is recognized in all 50 states. The Pharmacy Technician Certification Examination (PTCE) produced by the PTCB adheres to the Uniform Guidelines on Employee Selection adopted by the Equal Employment Opportunity Commission (EEOC), American Disabilities Act (ADA), Civil Service Commission (CSC), Department of Labor (DOL), and Department of Justice (DOJ). PTCB procedures also adhere to the Professional Examination Service's (PES) guidelines for the Development, Use, and Evaluation of Licensure and Certification Programs (1995). Additionally, PTCB's pharmacy technician certification examination is accredited by the National Commission for Certifying Agencies (NCCA).

The proposal submitted by the PTCB includes much strength with few weaknesses. The strengths of the proposal particularly salient to this RFQ include the following:

- A management system and team that is deep in certification/licensing, information technology, project management, law, and accounting.
- An organizational infrastructure (e.g., management and technical staff and a Ph.D. Psychometrician with credentials from a high-quality program) that provides evidence that PTCB will be able to fulfill the obligations outlined in this RFQ if awarded.
- Strict adherence to standards for certification/licensure testing (and testing practice more generally) as detailed by the American Educational Research Association, American Psychological Association, National Council on Measurement in Education.
- Working with professionals from American College Testing (ACT) and a diverse group of subject matter experts and job analysis panelists with an average of 25 years of experience on the job analysis study specifications, construction of survey and overall design and conduct.
- Clear and comprehensive reporting of the sampling plan (frame), analytic sample used for the job analysis study.
- Psychometrically rigorous approach to analyzing the ratings from the job analysis survey (e.g., a nested approach that provided a natural way to combine the three survey subscales into a single measure/score; item response theory (Rasch model) for scaling the responses).
- A recently conducted job analysis (2012) in conjunction with ACT.

- Comprehensive test blueprint that appropriately links KSAs with tasks; descriptions of decisions made by the SMEs to change and/or delete KSAs from the test blueprint.
- Item development directed/supervised by appropriately trained employees.
- Thorough and rigorous approach to item piloting, field-testing, and performance evaluation using classical test theory (CTT) and item response theory (IRT) approaches.
- Used appropriate techniques for examining differential item functioning (DIF) or “item bias” to ensure that test items are not unfair (psychometrically and culturally) to the various groups being tested.
- Appropriate use of standard setting technique for establishing the cut score for passing the exam; clear explanation of masters and non-masters relative to decision rules (i.e. Type I and Type II errors).
- Item response theory based equating of different test forms; this also facilitates interpreting the cut-score across different test forms.
- Thorough and comprehensive item anchoring strategy to facilitate new or parallel test forms using IRT.
- Use of both classical test theory (IRT) and item response theory (IRT) in evaluating item performance (very thorough on this point), test scoring, equating, and evaluating traditional score reliability and IRT-based conditional errors of measurement along the score scale.
- Calculation of classification (i.e. decision) indices based on pass/fail scores.
- Strong evidence of item bank security (physical, system and security).
- Item bank quality assurance procedures are rigorous.
- Strong evidence of high-quality standardization of test administration/delivery.
- Data storage/security procedures are comprehensive and state-of-the art.

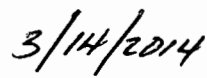
### *Weakness*

### *Stability Reliability*

Regarding stability reliability, the authors state their case for not deriving stability score reliability for the test because candidates’ knowledge does not (or is not expected to) remain stable over time. Indeed this is one reason not to calculate and report stability reliability; however, it may be advantageous to do so. A study of the efficacy of this type of score reliability is probably warranted.

Respectfully submitted,

  
Larry R. Price, Ph.D.

  
Date

## References

- American Educational Research Association/American Psychological Association/National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Orlando, FL: Harcourt, Brace, Jovanovich Publishers.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to Measurement Theory*. Belmont, CA: Brooks/Cole Publishers.
- Kline, T. J. B. (2005). *Psychological Testing*. Thousand Oaks, CA: Sage Publications.
- Kolen, M. J., & Brennan, R. L. (1995). *Test Equating: Methods and Practices*. New York, NY: Springer-Verlag Publishers.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory, 4th ed.* New York, NY: McGraw-Hill Publishers.
- Cizek, G. J. (2001). *Setting Performance Standards: Concepts, Methods, and Perspectives*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Cohen, R. J., & Swerdlick, M. E. (2010). *Psychological Testing and Assessment: An Introduction to Tests and Measurements, 7th ed.* New York, NY: McGraw-Hill Publishers.
- Haladyna, T. M. (2004). *Developing and Validating Multiple-Choice Test Items, 3rd ed.* Mahwah, NJ: Lawrence Erlbaum Publishers.
- Czaja, R., & Blair, J. (2005). *Designing Surveys: A Guide to Decisions and Procedures*. Thousand Oaks, CA: Sage Publications.
- Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E., & Tourangeam, R. (2009). *Survey Methodology, 2nd ed.* New York, NY: Wiley Publishers.
- de Alaya, R. J. (2009). *The Theory and Practice of Item Response Theory*. New York, NY: Guilford Publishers.

Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. (2007). Standardized Conditional SEM: A Case for Conditional Reliability. *Applied Psychological Measurement*, 31, 3, 169 – 180.

## OVERALL EVALUATIONS SUMMARY

EVALUATOR #1

TEST SERVICE	DESIGN	DEVELOPMENT	IMPLEMENTATION	DEFENSIBILITY	TOTAL
NHA	132	193	191	60	576
PTCB	140	195	197	60	592

EVALUATOR #2

TEST SERVICE	DESIGN	DEVELOPMENT	IMPLEMENTATION	DEFENSIBILITY	TOTAL
NHA	140	169	165	56	530
PTCB	130	191	173	60	554

## TOTALS

TEST SERVICE	DESIGN	DEVELOPMENT	IMPLEMENTATION	DEFENSIBILITY	GRAND TOTALS
NHA	272	362	356	116	1106
PTCB	270	386	370	120	1146