SAMPLE FUNDED RESEARCH PROPOSAL

Research Proposal to the
Joint Research Committee – NCSBN

1. PROJECT TITLE:
An Investigation of Item Response Time Distributions as Indicators of Compromised NCLEX Item Pools

2. PRINCIPAL INVESTIGATOR:		CO-PRINCIPAL INVESTIGATOR: (if applicable)
Name:	Steven L. Wise	G. Gage Kingsbury
Title:	Professor of Graduate Psychology	Director of Research
Institution:	James Madison University	Northwest Evaluation Association
Address:	821 S. Main St. MSC 6806	5885 SW Meadows Road Suite 200
City, State and ZIP	Harrisonburg, VA 22807	Lake Oswego, OR 97035
Phone Number:	540-568-6706	503-624-1951
Fax Number:	540-568-7878	503-639-7873
E-mail Address:	wisesl@jmu.edu	

3. RESEARCH FUNDS REQUESTED:

Total: $ 17,500

4. PROPOSED PROJECT DURATION:

Starting Date: August 2005
Ending Date: May 2006

5. BRIEF ABSTRACT OF THE PROJECT: (*No more than 200 words*)

Item response times can provide valuable information for assessing the degree to which an item pool has been compromised. Advance knowledge will tend to decrease the amount of time an examinee spends on the item. Increased advance knowledge should result in higher accuracy for rapid responses.

We propose to conduct a study to explore NCLEX data for evidence of shifts in response time distributions and accuracy of rapid responses. We will compare pilot test data to data from a single, later time point using three sets of candidates: 10,000 US-educated first time test takers, 10,000 non-US educated first time test takers, and 10,000 non first time test takers. Item responses, item identifiers, final score, and item response times will be needed for each candidate.

Item responses that show rapid correct responding will be identified and accumulated. For any items for which rapid responding is more common that anticipated, the item will be checked against other item characteristics from field test results. The goal is to develop a procedure that measurement practitioners can use to assess the degree to which advance knowledge has compromised their CAT pools. The NCLEX program can then make more effective decisions regarding item pool rotation.

An Investigation of Item Response Time Distributions as Indicators of Compromised NCLEX Item Pools

Project Summary

Issues Addressed and Importance
Item response times can provide valuable information for assessing the degree to which an item pool has been compromised. Advance knowledge will tend to decrease the amount of time an examinee spends on the item. Increased advance knowledge should result in higher accuracy for rapid responses.

Methodology
We propose to conduct a study to explore NCLEX data for evidence of shifts in response time distributions and accuracy of rapid responses. We will compare pilot test data to data from a single, later time point using three sets of candidates: 10,000 US-educated first time test takers, 10,000 non-US educated first time test takers, and 10,000 non first time test takers. Item responses, item identifiers, final score, and item response times will be needed for each candidate.

Intended Outcomes and Importance for the NCLEX Program
Item responses that show rapid correct responding will be identified and accumulated. For any items for which rapid responding is more common that anticipated, the item will be checked against other item characteristics from field test results. The goal is to develop a procedure that measurement practitioners can use to assess the degree to which advance knowledge has compromised their CAT pools. The NCLEX program can then make more effective decisions regarding item pool rotation.

Proposal Narrative

Maintaining test item pool security is a challenge common to all high-stakes CAT programs. The validity of inferences made on the basis of examinee test scores is dependent on the accuracy and stability of the item pool's IRT parameters. However, in high-stakes programs, such as the NCLEX, there are continual attempts by individuals or test preparation organizations to acquire advance knowledge of items in the pool. And to the degree to which items are known by examinees prior to taking their CATs, they will become easier and their difficulty parameters will not be appropriate for estimating examinee proficiency. Thus, advance knowledge of items represents a serious threat to score validity. Such advance knowledge is commonly obtained as items are exposed to examinees, and those examinees pass item content information on to other individuals or organizations.

The NCLEX-RN and NCLEX-PN programs have dealt with this problem by frequently changing item pools, which minimizes the exposure of individual items. This strategy, however, greatly increases the resources required to maintain score validity, as many new items need to be developed. Moreover, it is not easy for measurement practitioners to assess the degree to which a given item pool has been compromised and thereby judge whether scores have been compromised between pool changes. If the pool is changed too frequently, an excessive amount of resources will be devoted to item development. In contrast, if the pool is not changed frequently enough, test score validity will be threatened.

Item response times can provide valuable information for assessing the degree to which an item pool has been compromised and this response time information is routinely collected during an NCLEX administration. This information has been used to examine candidate behavior for individuals who run out of time, but hasn't yet been exploited to examine item exposure.

Item response time has been previously shown to be useful in identifying unusual examinee behavior. Schnipke and Scrams (2002) showed that at the end of speeded, high-stakes computer-based tests (CBTs), some examinees strategically switch from trying to identify correct answers to the items (termed *solution behavior*) to very rapidly submitting answers before time expires (termed *rapid-guessing behavior*). Schnipke and Scrams found that rapid-guessing behavior yielded answers that were essentially random, and therefore provided little information regarding examinee proficiency. Similarly, Wise and Kong (2005) found that rapid-guessing behaviors frequently occur during unspeeded low-stakes CBTs. They showed that, in a low-stakes context, rapid-guessing behavior indicates a lack of examinee test-taking effort, as shown by examinees responding before they had time to read and comprehend an item.

In both of the studies described above, rapid-guessing behavior was exhibited by examinees who did not try to solve the challenge posed by a particular test item because they either did not have time or did not feel like trying. There is, however, another reason that an examinee might answer quickly—if he or she had advance knowledge of the item (and presumably, its correct answer). We might more accurately term such an occurrence *rapid-choice behavior*, because the response represents a purposeful choice rather than a guess. Rapid-choice behavior due to advance knowledge of the item can be differentiated from rapid-guessing behavior by examining the accuracy of the answer provided. Rapid-guessing behavior will yield responses whose accuracy is close to that expected by chance (Schnipke & Scrams, 2002; Wise & Kong, 2005), while rapid-choice behavior should yield responses with much higher accuracy. Thus, rapid-choice behaviors would ideally be characterized as quick, accurate responses.

In practice, however, it is unrealistic to expect that any advance knowledge an examinee might have will always (or even typically) be complete. It is more reasonable to assume that some examinees will have only partial knowledge of an item (i.e., knowing in advance either some of the item text or the task being asked by the item)—enough to make the item easier for these examinees, but not necessarily enough to yield rapid-choice behavior. Any procedure designed to detect compromised item pools should therefore be sensitive to both partial and complete advance knowledge.

Proposed Study

We believe that, in general, advance knowledge of an item will affect the frequency distribution of response times for that item. Specifically, advance knowledge (either partial or complete) will tend to decrease the amount of time an examinee spends on the item. This suggests that the greater the proportion of examinees with advance knowledge, or the more complete the advance knowledge tends to be, the greater the degree to which the distribution of response times will be affected.

In addition to the effects of advance knowledge on an item's response time distribution, there should be an accompanying effect on the accuracy of relatively rapid responses to

the item. That is, increased advance knowledge should result in higher accuracy for rapid responses. The greater the degree of advance knowledge for an item, the greater the increase in accuracy will be observed.

We propose to conduct a study in which we will explore NCLEX data for evidence of shifts in response time distributions and accuracy of rapid responses. The response time distributions and accuracy rates for a set of pilot tested items will serve as reference distributions. Then, after these items have been in the operational pool for a period of time, a new set of response time distributions and accuracy rates will be generated for the same set of items based only on the later NCLEX administrations. If examinees have gained advance knowledge of these items, comparisons between the data from the two time points should reveal evidence of the shifts predicted above.

In this initial study, we will compare the pilot test data to the data from a single later time point. If the predicted shifts in response time distributions and accuracy rates are observed, then we will propose additional studies to refine our research methods. The ultimate goal is to develop a procedure that measurement practitioners can use to assess the degree to which their CAT pools have been compromised by advance knowledge. This will allow them to make more effective decisions regarding when item pools need to be changed.

Two elements differentiate this study from the work NCSBN is currently doing with Caveon. First, the approach used here is designed for use with adaptive tests, which should enable it to identify more fine-grained deviations from expectation. Second, the methodology used in this series of studies will become available to NCSBN without additional cost.

Design

A set of 500 items, calibrated within the past 3 calendar years and in active use in the NCLEX-RN test during spring of 2004 will be selected for use. Item difficulties and item identifiers from field testing will be needed, as will the frequency distributions of each item's response times. Based on this information, thresholds for identifying rapid response will be generated for each item.

Three sets of candidates will be identified who were tested in spring of 2004. The first set will consist of 10,000 US-educated first time test takers. The second set will be an equal number of non-US educated first time test takers. The third set will consist of an equal sized set of non first time test takers. Item responses (correct/incorrect), item identifiers, final score, and item response times will be needed for each candidate.

For each set of candidates and for each item in the set, item responses that are evidence of rapid correct responding will be identified and accumulated. For any items for which rapid responding is more common that anticipated, the item will be analyzed to identify whether other item characteristics, including item difficulty and item fit differ from field test results.

If the first study indicates that the methodology is useful in identifying oddly-performing items, a second study will be proposed which will evaluate a process for continuous evaluation of rapid responding as a method for identifying item exposure. This second

study will compare rapid responding to recalibration to percent-correct drift as processes for item identification.

References

Schnipke, D. L., & Scrams, D. J. (2002). Exploring issues of examinee behavior: Insights gained from response-time analyses. In Mills, C. N., Potenza, M.T., Fremer, J. J., & Ward, W. C. (Eds.). *Computer-based testing: Building the foundation for future assessments.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18,* 163-183.

Project Timeline

The project analyses will be completed and a report will be submitted to NCSBN within nine months of project approval. A progress report will be presented at the Spring meeting of the JRC, and a final written report will be submitted at the end of the project (May 2006).

BUDGET SUMMARY
Research Proposal to the
Joint Research Committee – NCSBN

BUDGET SUMMARY

CATEGORY	**Cost**	*Use these additional columns to split out costs for each phase, if proposing a multi-phased project*			
DIRECT COSTS*					
1. Personnel salaries and wages	$17,000				
2. Fringe benefits					
3. Consultants and contracts					
4. Travel					
5. Supplies and materials	$500				
6. Communications (Telephone, postage, etc.)					
7. Equipment (Purchase)					
8. Other (Equipment rental, etc.)					
TOTAL	$17,500				

**Note that Indirect costs or other overhead charges will not be reimbursed*

BUDGET JUSTIFICATION NARRATIVE

Project personnel costs include data analysis and report writing. Funds for supplies will cover expenses for printing, copying, postage, and long-distance telephone calls.