Draft: 06-03-04

FEASIBILITY AND BARRIERS FOR A NATIONAL ITEM BANK: IF WE BUILD IT WILL THEY COME?

Dennis A. Revicki, Ph.D.

Center for Health Outcomes Research, MEDTAP International, Bethesda, Maryland.

For more information contact: Dennis Revicki, Ph.D., Director, Center for Health Outcomes Research, MEDTAP International, 7101 Wisconsin Ave., Suite 600, Bethesda, MD 20814. Voice: 301-654-9729; Fax: 301-954-9864; E-mail: revicki@medtap.com.

INTRODUCTION

The development and application of item response theory (IRT) and computer adaptive testing (CAT) methods to health status outcomes have the potential to provide efficient collection of better quality, more precise self-reported data from patients participating in clinical trials, clinical observational studies, and in population surveys both nationally and in managed care organizations (McHorney, 1997; Revicki & Cella, 1997). CAT allows the more efficient data collection and enables the application of IRT-based assessment. Although IRT and CAT methods are currently under evaluation in a number of different disease areas and for assessing different domains of patient-reported outcomes (PROs), they have not been widely applied in clinical or health services research. The techniques of IRT, item banking, and CAT provide an innovative solution to the challenges of assessing health status in subjects with varying problems across different health domains.

It is now technically feasible and possible to develop and make available a national item bank for the implementation of IRT-based, tailored testing; but much research and measurement work remains to bring these ideas into practical application (McHorney, 2002; Bode et al., 2003). Even if sufficient financial support was available for accomplishing a viable national item bank, including software for administering and scoring IRT-based CATs, it is uncertain whether or not instrument developers, clinicians and clinical researchers, the pharmaceutical or other health care industries, and regulatory agencies would fully accept this measurement approach.

FEASIBILITY AND ACCEPTANCE OF A NATIONAL ITEM BANK

There are significant challenges associated with developing, evaluating and maintaining a national item bank. However, one of the largest challenges may be associated

with the practical application of this measurement approach once the item bank and measurement software is developed and tested by various stakeholder groups. These stakeholders, such as instrument developers, clinical researchers, the pharmaceutical industry, and federal regulatory authorities may resist acceptance of these new measures. In this next section, several potential barriers associated with each of these stakeholder groups will be discussed.

Instrument Developers

In the current situation, instrument developers may have a vested interest in maintaining the integrity of the existing generic- and disease-specific measures that they have developed over time that include static, fixed item sets for all respondents. For some health outcomes researchers, their academic reputations have been built upon developing and evaluating their health-related measures, and this research represents a significant contribution to understanding the impact of disease and treatment from the patient's perspective. Although several instrument developers are working at developing item banks, IRT models and CATs based on their assembled items from various health status instruments, it is uncertain how ownership of these new IRT-based measures will be handled.

Some developers market and license these health measures and receive from a modest to a relatively substantial revenue stream from their instruments. What is the incentive for instrument developers to contribute their copyrighted items to the larger national item bank? In addition, if taken to its logical conclusion, if a national item bank is available and researchers actively use this bank and related CAT methods, there is no real need for 'brand name' instruments. Their may be considerable resistance to cooperating in any national

measurement effort and those researchers who do collaborate in the development and evaluation of the national item bank may find themselves in a monopoly position.

There is no easy way to address this potential barrier, since there are important academic (i.e., promotion, tenure) and financial stakes at risk. Possible solutions include involving a broad range of researchers in the development and psychometric evaluation process, since there will need to be extensive field studies required to build, test and maintain the bank. Second, a modest royalty (0.25 to 1 cent) might be offered to instrument developers associated with each time one of their items is used in a study. However, this would require some independent entity to manage the national item bank and some sort of new financial model for deriving revenues.

Clinical Researchers

Although many may recognize that an approach is needed for measuring health status across the relevant continuum of functioning and well-being for longitudinal, chronic disease studies where subjects may deteriorate at different rates over time, they may not accept IRT-based tailored tests as the best approach. A static health status instrument may not be able to capture these effects for long-term studies. It is uncertain whether clinical researchers will accept and include the IRT based measures in studies. First, significant barriers are associated with their understanding and unfamiliarity with the IRT/CAT method, and continued skepticism from some clinicians about PROs, in general, may limit acceptance. Second, application of CAT will require computer-administered instruments and there may be practical feasibility or budget issues associated with applications in studies. The most significant barrier related to practicality, given that desktop or laptop computers may not be very practical in the clinical context. This problem may be addressed by deriving 'static'

IRT-based instruments using the item bank, tailored to the relevant patient population and study or through the use of handheld computers for administering health status instruments.

Food and Drug Administration

The Food and Drug Administration (FDA) (and other regulatory agencies) have recently accepted the importance of PROs for understanding effectiveness of treatments from the patient's perspective. Although there has been increased understanding about PRO measures and methods (Burke, 2001; Revicki, in press), regulatory agencies have few staff members who really understand psychometrics, and none who understand IRT and CAT. Currently, the FDA focuses on static, disease-specific health status and other PRO measures, and requests information on instrument development, face and content validity, and measurement characteristics (i.e., reliability, validity, responsiveness) as part of the documentation package underlying labeling or promotional claims for PRO benefits for all products. It may be difficult, due to institutional culture and limited psychometric training, for the FDA staff to understand and accept measures where the item content is different across individual study subjects and that may vary in item content within subjects over the duration of the study. There may be some cognitive discontinuity in adjusting from a situation where the focus is on static instruments to one where outcomes researchers endorse no set instrument but banks of items designed to measure different domains. There are also challenges associated with meeting the requirements for validated documentation of clinical trial data when measures are collected through electronic data capture. But, these technical challenges are likely to be more surmountable than the cognitive ones.

Pharmaceutical Industry

To a great extent, the FDA drives the actions of the pharmaceutical industry. Few pharmaceutical companies will select PRO measures that will not be acceptable to the FDA. There may be reluctance for industry researchers to take a chance that the regulatory agencies will accept CAT and IRT based assessments for evaluating the patient-related effectiveness of new products. There will need to be continued dialogue and exchange between health outcomes research, psychometricians, the FDA and industry researchers to determine the best and most scientifically acceptable way to incorporate IRT based measures and CAT into clinical trial programs. The start may be with more tailored tests for different patient populations and applications that are more consistent with static measures used currently. Regulatory agencies are interested in PRO measures that provide assessment of the effectiveness of new treatments and, based on the current situation require evidence supporting content validity, good psychometric characteristics and guidance on interpretation of results. To the extent that industry can provide this documentation, the CAT-based measures may prove acceptable for drug evaluations. The key issue for industry relates to what are the advantages of using CATs in evaluating the effectiveness of new treatments? For industry, the additional expense associated with application of CAT in clinical trials must result in some positive trade-off in the increased probability for demonstrating effectiveness of the new treatment.

FINANCIAL CONSIDERATIONS IN DEVELOPMENT AND MAINTAINING A NATIONAL ITEM BANK

There are significant challenges associated with financial support for a national item bank, and some type of publically and privately funded entity may be the best solution. The most important aspects of this entity will be to balance providing ready access, IRT findings,

and any developed software for administering CAT, with having enough revenues to remain viable and continue to update and maintain the item bank. However, without the cooperation and support of Federal agencies, clinical researchers, instrument developers, the pharmaceutical industry and managed care organizations, it is unlikely to survive. Several financial models exist, such as charging modest user fees for use of item bank measures and software and/or royalties. A useful model to consider is that of the Educational Testing Service in developing and supplying college entrance examinations, professional competency tests and other achievement tests. However, at this time it is uncertain whether there is sufficient demand for these type services for PRO assessment.

SUMMARY AND CONCLUSIONS

If we (or they) build it, will they come? There are considerable challenges associated with building and maintaining a national item bank and it is uncertain whether there is sufficient interest among key stakeholders for IRT-based and CAT measures. The most convincing activity is demonstrating that the approach is feasible, psychometrically sound and useful in a specific application. Demonstrated success opens up the possibility of more widespread acceptability and application. As part of the development effort, there needs to be continued meetings and discussion with psychometricians, instrument developers, clinical researchers, the FDA, pharmaceutical industry researchers, and a managed care organizations about the advantages and disadvantages of a national item bank. These discussions may be as important as the investment in the actual development of a national item bank and CAT software.

REFERENCES

Bode RK, Lai JS, Cella D, Heinemann AW. Issues in the development of an item bank. Arch Phys Med Rehabil 2003;84:S52-S60.

Burke LB. US regulation of pharmaceutical outcomes research. Value in Health 2001;4:5-7.

McHorney CA. Generic health measurement: past accomplishments and measurement paradigm for the 21st century. Ann Intern Med 1997;127:743-750.

McHorney CA. Use of item response theory to link 3 modules of functional status items from the Asset and Health Dynamics Among the Oldest old study. Arch Phys Med Rehabil 2002;83:383-394.

Revicki DA. Use of health-related quality of life measures by industry and regulatory agencies in evaluating oncology therapies. In Lipscomb J, Gotay C, Snyder C. (eds.). Outcomes Assessment in Cancer. Cambridge: Cambridge University Press, in press.

Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. Qual Life Res 1997;6:595-600.