The articles in this supplement to *Health Education Research* grew out of discussions among the coeditors as to the best way to further promote the use of advanced psychometric methods in health education and health behavior research. Item response modeling (IRM), also referred to as item response theory, is well established and widely implemented in educational measurement, but its application is lagging in our area. A number of issues seem to stunt the application of IRM methods: (i) few IRM applications have been presented in the context of health education and health behavior research; (ii) lack of awareness as to what IRM can do beyond assessing the psychometric properties of a scale; (iii) lack of psychometricians trained in our field; (iv) current software is not user friendly and (v) few training opportunities are available to researchers in our area. In an attempt to address some of these issues, the National Cancer Institute (NCI) in collaboration with the coeditors of this supplement devised strategies to promote the development of better scales through improved methods. Although this supplement is a key product of the activities undertaken by the coeditors, NCI has a history for providing support to the development of improved methods, including IRM. In September 1999, the Agency for Healthcare Research and Quality and the NCI jointly sponsored a symposium aimed at improving health outcomes methods and devoted a significant amount of time to explaining IRM and the usefulness of IRM-computerized adaptive testing methods for assessing health outcomes (papers were published in a supplement to *Medical Care* in 2000) [1]. In July 2004, NCI co-sponsored a conference that promoted the use of IRM applications for health outcomes assessment as well as focusing on the usefulness of the methods for health education and health behavior assessment (http://outcomes.cancer.gov/conference/irt/). In planning for the conference, it became apparent that the method was more widely used in health outcomes assessment, which may partly be explained by earlier efforts. Realizing the gap in health education and health behavior research, the coeditors were prompted to work on this supplement. Given the dual focus of the 2004 measurement conference, the health outcomes assessment papers are scheduled to be published in a supplement to the *Quality of Life Research* journal and the health education and health behavior papers are appearing in this supplement. The focus of these two supplements differs. The health outcomes supplement focuses on issues associated with managing an item bank and applications of computerized adaptive testing. Although important, it appeared premature to address these issues here, given the lack of understanding of

---

1Department of Pediatrics, Centre for Community Child Health Research, University of British Columbia, L408, 4480 Oak Street, Vancouver, BC V6H 3V4 Canada, 2Graduate School of Education, University of California, Berkeley, CA 94720, USA, 3Children’s Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, 1100 Bates Street, Houston, TX 77030, USA and 4National Cancer Institute, Division of Cancer Control and Population Sciences, Behavioral Research Program, Health Promotion Research Branch, Bethesda, MD 20892-7335, USA

*Correspondence to: L. C. Mâsse.
E-mail: lmasse@cw.bc.ca
the basics of IRM in our field. The papers presented in this supplement provide a comprehensive look at the usefulness of IRM to address specific, but common, measurement issues encountered in health education and health behavior research.

In addition to these initiatives, NCI sponsored a workshop and symposium in July 2006 at the International Society of Behavioral Nutrition and Physical Activity to promote the application and understanding of these methods. Understanding the need to improve current IRM software, in 2004 NCI issued a Small Business Innovation Research and Small Business Technology Transfer program announcement to improve the development of IRM software for outcomes and behavioral assessment (http://grants2.nih.gov/grants/guide/notice-files/NOT-OD-04-055.html). Finally, NCI is a major partner of the National Institutes of Health (NIH) cooperative initiative Patient-Reported Outcomes Measurement Information System (http://www.NIHpromis.org) aimed at developing an item bank repository that is publicly available and which can serve as a template for computerized adaptive testing for health outcomes assessment. It is hoped that this supplement as well as the other ongoing initiatives will lead to improved methods in health education and health behavior research.

The topics of the papers presented in this supplement as well as the data sets used as examples of these IRM applications were carefully selected to represent the maximum level of interest to a broad range of readers. Data sets and scales that are widely used in health promotion and health behavior research were selected; hence, many of the applications used the behavioral change consortium (BCC) data sets [2]. The BCC was an initiative funded by NIH in 1998, which included 16 projects and >100 researchers. The focus of these consortium projects was to stimulate innovative strategies to achieve and maintain long-term behavior change as it relates to tobacco use, insufficient exercise, poor diet and alcohol abuse with a particular focus on understanding the mechanisms of behavior change. Scales selected from the BCC initiative were those that are widely used in the field and across many disciplines. However, applications in this supplement do not exclusively use the BCC data sets as the aim of this supplement is to present issues which would be of greatest interest to the readers.

This supplement includes nine papers which cover a range of IRM applications. A premise of IRM is that we can enhance our measures the more we understand how they work in different groups and different contexts. A major emphasis is on providing new tools for exploring scale and item functioning, based on fitting a latent variable to the items. The first two papers by Wilson et al. introduce the IRM approach [3, 4]. They take an applied approach to present these concepts and utilize the widely used self-efficacy scale to explain the fundamentals of IRM. After discussing the basic ideas behind the approach, the first paper describes some typical outcomes of analysis using item response models. The second paper is devoted to helping those who are more familiar with classical test theory see what extra results are available in the newer approach. These papers are followed by three IRM applications to assess the psychometric properties of a scale using IRM [5–7]. The paper by Mäße et al. compares the results obtained with classical test theory, IRM and confirmatory factor analysis, and clearly demonstrates that IRM provides complementary information not available by other methods. The papers by Watson et al. and Heesch et al. present valuable information for improving old scales, and highlight that many of our scales do not have enough items to detect change over time. As many of our scales are used to detect the efficacy/effectiveness of behavioral interventions, the scales need enough power to detect change over time. The last four papers present advanced uses of IRM. While IRM is a powerful method that provides useful information about the psychometric properties of a scale, its advanced uses hold even more promise. The paper by Allen and Wilson [8] introduces multi-dimensional IRM and shows that analyzing the subscales in a multi-dimensional manner increased the power of these scales. The impact on the reliability estimate makes a convincing argument that using more complex methods can improve the power of our scales to detect change over time. The paper by Watson et al. [9] also uses
multi-dimensional IRM, but for correcting for social desirability of response bias, and demonstrates how doing so increased the correlation with a behavioral outcome. The paper by Baranowski et al. [10] used IRM’s differential item functioning to assess if the interpretation of the items changed as a result of an intervention or from being exposed to the measurement instrument. Although minimal bias was observed, it is crucial that we understand the stability of our scales and determine if their properties change as a result of being exposed to an intervention or after being exposed to the measurement instrument. The last paper, by Masse et al. [11], introduces equating methods to compare different studies that have employed the same scales but included different items. As there is still much variability in the items selected, the equating method can serve to facilitate comparison of studies that uses different items to measure a common construct. In summary, this supplement presents a number of IRM applications. These applications should increase understanding of the usefulness of IRM and ultimately lead to better measures. We believe this supplement is timely and that it provides a primer to IRM applications which can serve as a stepping-stone to promote further applications of the method. The papers included in this supplement were collectively chosen to present certain IRM applications. Many important IRM applications have not been presented (such as the development of an item bank, computerized adaptive testing and measurement of change). However, this supplement provides numerous applications which can have widespread use in the field of health education and health behavior research. We hope that these applications will advance the field by enhancing the quality of the measures available and the appropriateness of the methods involved.

Acknowledgements

This supplemental issue was sponsored by the NCI. L.C.M was at the NCI, Health Promotion Research Branch, Behavioral Research Program when the supplemental issue started. The views presented in this paper and the supplement represent those of the authors and not those of the NCI or any of the employers’ authors.

References