HEALTH EDUCATION RESEARCH
Theory & Practice

Vol.21 (Supplement 1) 2006
Pages i73–i84
Advance Access publication 3 October 2006

Introducing multidimensional item response modeling in health behavior and health education research

Diane D. Allen and Mark Wilson*

Abstract

When measuring participant-reported attitudes and outcomes in the behavioral sciences, there are many instances when the common measurement assumption of unidimensionality does not hold. In these cases, the application of a multidimensional measurement model is both technically appropriate and potentially advantageous in substance. In this paper, we illustrate the usefulness of a multidimensional approach to measurement using an empirical example taken from the Behavior Change Consortium. Data from the Treatment Self-Regulation Questionnaire have been analyzed to investigate whether self-regulation can be regarded as a single construct, or if it has multiple dimensions based on the type of regulation or motivation that participants say helps them consider an improvement in healthy behavior. Comparison with consecutive analyses shows the advantages of multidimensional measurement for interpreting participant-reported data.

The question of dimensionality

Health education and behavioral change programs directed toward smoking cessation, diet improvement or maintenance of adequate physical activity involve complex mediators of behavior change. To assess the effectiveness of such programs, researchers will need to consider that some of these mediators may be multidimensional (i.e. have more than one dimension). In creating scales or surveys to examine such mediators, researchers want to have sufficient items for adequate reliability (i.e. a high proportion of the variance considered true rather than attributable to error) but not so many that they overburden participants. Interpreting a single score across all items (i.e. conducting a composite analysis) may result in adequate reliability at the expense of loss of information about the separate dimensions. Interpreting separate scores on each of the dimensions (i.e. conducting a consecutive analysis) may improve information about each subscale, but at the expense of lower reliability for the subscales than for the composite scale. One solution to this dilemma, and the focus of this paper, is to conduct a simultaneous or 'multidimensional' analysis of the subscales.

One proposed mediator of behavior change is examined with the Treatment Self-Regulation Questionnaire (TSRQ) [1]. Figure 1 illustrates a way to understand the structure of this scale. Subdividing a construct like self-determination can potentially continue until each item represents a single dimension. However, the number of dimensions depends on both the substantive meaning of these subscales and the number of items needed to define them with sufficient reliability. One of the authors of the TSRQ confirmed for us the substantive subscales shown in Fig. 1 for this scale and the items associated with each subscale (G. C. Williams, personal communication, 26 August 2004).

Graduate School of Education, University of California, Berkeley, CA 94720, USA
***Correspondence to: M. Wilson.**
E-mail: markw@calmail.berkeley.edu

doi:10.1093/her/cyl086

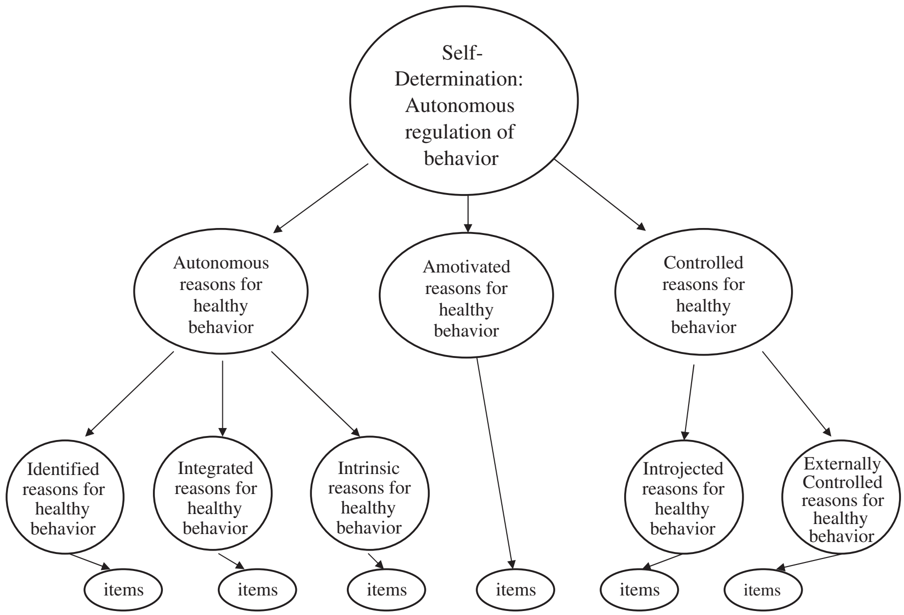

Fig. 1. Example of a multidimensional instrument: the TSRQ and its reasons for healthy behavior.

The number of items and the statistical model employed to analyze data resulting from a scale like the TSRQ can have an impact on the reliability estimate and the interpretation of its subscales. In this paper, we consider three methods to analyze the TSRQ: (i) A 'composite' analysis examines self-determination based on a single score incorporating all items across subscales for each participant; (ii) A 'consecutive' set of analyses examines each subscale alone as a separate subscale of reasons for healthy behavior and analyzes data from these subscales one at a time; and (iii) A 'multidimensional' analysis examines the different subscales simultaneously, incorporating the responses to all the items indicating reasons for healthy behavior on the whole scale and capitalizing on the correlations between subscales. The actual use of the TSRQ most closely follows the consecutive method with average scores on subscales reported separately. The composite method is added somewhat artificially here for illustrative purposes since it is commonly employed with many scales used in health education research and the behavioral sciences.

The purpose of this paper is to illustrate the usefulness of a multidimensional approach to measurement by comparing these three different analyses of the TSRQ [1]. We use data from the Behavior Change Consortium (BCC) which was a group of 15 projects designed to explore how people manage changes in health behaviors [2]. Although several of the BCC projects used the TSRQ, this paper targeted data from a project promoting diet improvement and smoking cessation. We first describe the TSRQ and the context of our data within the BCC project. The following section briefly introduces and describes a multidimensional version of the item response model described in other papers in this supplement [3, 4]. This is followed by a summary of results obtained using the multidimensional model, and comparisons with consecutive and composite analyses of the data arising from the TSRQ.

The TSRQ

Self-determination, the construct underlying the TSRQ, means one chooses behavior for one's own reasons rather than basing choices on other people's expectations or desires. It is defined as deciding on one's actions independent of others. People who are self-determined feel free to do what is interesting, personally important and vitalizing (http://www.scp.rochester.edu/SDT/). Evidence for Self-Determination Theory (SDT) as applied to choices for healthy living would include indications that greater self-determination or 'self-regulation' is associated with an increase in and maintenance of healthy behaviors [5]. Our data set, however, only included pre-intervention data.

The TSRQ assesses self-determination by asking how people regulate their behaviors regarding health, whether the behavior in question is diet improvement or smoking cessation [1]. Self-regulation is thought to be a continuum anchored by 'autonomous' or self-determined reasons for behavior at one end and 'controlled' reasons for behavior at the other [5]. Autonomous behavior 'has an internal perceived locus of causality and is experienced as chosen and volitional' [6, p. 768]; autonomous items are divided into identified (ID), integrated (INT) and intrinsically motivated (INTR) reasons for healthy behavior. Controlled behavior has an external causality, such that individuals feel pressured by demands and contingencies; controlled items are divided into introjected (IJ) and externally controlled (EC) reasons for healthy behavior. Amotivated (AMOT) behavior lacks an identified motive, e.g. 'I really don't think about it'. Thus, out of a single construct of self-determination, six distinct item sets evolved (see Fig. 1) that might then function as separate dimensions. Figure 2 depicts the items and their associated dimensions within the TSRQ.

The data from the TSRQ for this paper came from the University of Rochester BCC project. This project had two foci: (i) testing a SDT model of maintained smoking cessation and diet improvement and (ii) testing an intervention against a 'usual' treatment [5]. In the intervention, a counselor contacted participants four to six times to educate and promote autonomous motivation for change, and support for change if attempted. Smoking adults were recruited through physicians' offices and insurers, and through announcements and advertisements. Once recruited, participants were stratified by amount of low-density lipoprotein cholesterol in their lipid profiles. For those with higher cholesterol, diet counseling was incorporated along with counseling about smoking

The reason I would (eat a healthy diet/stop smoking) is:

1. **ID** Because I feel that I want to take responsibility for my own health.
2. **IJ** Because I would feel guilty or ashamed of myself if I did not eat a healthy diet/stop smoking.
3. **ID** Because I personally believe it is the best thing for my health.
4. **EC** Because others would be upset with me if I did not.
5. **AMOT** I really don't think about it.
6. **INTR** Because I have carefully thought about it and believe it is very important for many aspects of my life.
7. **IJ** Because I would feel bad about myself if I did not eat a healthy diet/stop smoking.
8. **INT** Because it is an important choice I really want to make.
9. **EC** Because I feel pressure from others to do so.
10. **AMOT** Because it is easier to do what I am told than think about it.
11. **INT** Because it is consistent with my life goals.
12. **EC** Because I want others to approve of me.
13. **ID** Because it is very important for being as healthy as possible.
14. **EC** Because I want others to see I can do it.
15. **AMOT** I really don't know why.

Choices offered:

Not at all			somewhat true			very true
1	2	3	4	5	6	7

Fig. 2. The TSRQ items and the dimensions to which the authors attribute them.

cessation. Participants completed the TSRQ for both diet improvement and smoking cessation; only baseline data were analyzed for this paper ($n = 931$).

The participants in this project completed 30 items [5]: 15 each related to diet improvement and smoking cessation (Fig. 2). The six dimensions were represented with different numbers of items: ID had six, INT had four, INTR had two, IJ had four, EC had eight and AMOT had six. Each item had seven response categories with the following designations: 1 = not at all, 4 = somewhat true, 7 = very true. Response categories were recoded 0–6 for ease of analysis. The characteristics of the participants who completed the TSRQ for diet improvement and smoking cessation are shown in Table I.

Table I. *Characteristics of the participants in this data set (n = 931)*

Participant characteristics	*n*	%[a]
Gender	928	
Male	333	36
Female	595	64
Race	928	
White, not Hispanic	765	82
Black, not Hispanic	121	13
Hispanic	20	2
Other	22	2
Age	930	
18–37	235	25
38–45	228	25
46–52	223	24
53–82	244	26
Marital status	929	
Not married	229	25
Presently married	355	38
Living with partner	87	9
Divorced	177	19
Separated	50	5
Widowed	31	3
Education level	781	
Less than high school diploma	68	9
High school diploma	228	29
Some college, trade, vocational or technical school	294	38
Four years college/graduated	88	11
Post-graduate work	103	13
Income (per year)	901	
$0–39 999	524	58
$40 000–79 999	290	32
$80 000+	87	10

[a]Percent values may not add to 100 for some characteristics because of rounding.

A multidimensional item response model

The potential usefulness of multidimensional item response models has been recognized for many years in the psychological and educational literature [7–11]. Development of these models is well-covered elsewhere [12, 13], including exploration of the consequences of looking at multidimensional data as if it is a single dimension [14]. Despite this development in the methodological literature, the application of multidimensional item response models in the behavioral sciences has been limited. This has probably been due (i) to the statistical problems that have been involved in fitting such models and (ii) to the difficulty associated with interpreting multidimensional item response models. The statistical problems of multidimensional item response models have been addressed with the development of the multidimensional random coefficients multinomial logit (MRCML) model [12]. The MRCML model is an extension of the Rasch family of item response models, and, in particular, is a direct extension of the model described elsewhere in this supplement by Wilson *et al.* [3]. It was developed to have sufficient flexibility to represent a wide range of Rasch family models, including those that apply to scales having either yes/no or Likert-type responses. The MRCML models can also apply to scales having multiple items arising from the same stem or more complex situations involving differential item functioning (e.g. see Baranowski *et al.*, this volume [15]), raters and other measurement features. The MRCML is built up from a basic conceptual building block. To illustrate this building block, assume that an Item i with ordered categories of response (e.g. strongly disagree, disagree, agree, strongly agree) indexed by k helps to measure a single unique dimension d ($d = 1, ..., D$). (If the item is intended to measure only one dimension, the model is labeled multidimensional 'between items'. If the item is intended to measure two or more dimensions, the model is labeled multidimensional 'within items'

[12]. The latter model is beyond the scope of this paper.) The relationship of this item and response to a participant's level of attitude on dimension d is depicted in Equation 1. Here, the probability that a participant's response is in category k of Item i (P_{ik}) rather than category $k-1$ (P_{ik-1}) is related to the level of that participant's attitude on that dimension (θ_d) and the relative difficulty of category k (δ_{ik}) to endorse with that level of agreement:

$$\log(P_{ik}/P_{ik-1}) = \theta_d - \delta_{ik}. \tag{1}$$

Moreover, each participant has several θ_d values, one for each dimension of attitude measured by the scale: $\boldsymbol{\theta} = (\theta_1, ..., \theta_D)$, where the dimensions are allowed to be non-orthogonal (i.e. correlated). For example, in the TSRQ, the dimensions as depicted in Fig. 1 are subcomponents of participant self-determination, allowing participants to have different levels of endorsement for 'autonomous', 'amotivated' or 'controlled' reasons for healthy behavior. Besides providing a way to model the several dimensions of the TSRQ, the MRCML models allow the researcher to determine whether participants use the response categories consistently across all items within a dimension (as in a 'rating scale model') or differently across different items (as in a 'partial credit model') [16].

The data for creating the MRCML model building blocks consist of the responses all the participants make to all the items in the scale. The researcher specifies which MRCML model to try, and then a complex estimation procedure—now greatly simplified by appropriate software—generates the numerical values for the items and participants on a log odds or 'logit' scale. These estimated values maximize the fit to the specified model. The estimated values include the item parameters and the population means, variances and covariances of the θ_D parameters as related to the scale. For this paper, we fit a partial credit MRCML model to the data, using marginal maximum likelihood (MML) estimation as implemented in the ConQuest software [17]. (The partial credit model fit better than the rating scale model using a likelihood ratio test, $\chi^2 = 521.4$, df = 145, $p < 0.0001$, so it was used throughout the analyses.) In this technique, log response probabilities are summed up over items and participants into a likelihood function. Maximum likelihood estimates and asymptotic standard errors are found iteratively using the alternating estimation-maximization algorithm, and person estimates are calculated using an Expected A Posteriori (EAP) approach [18, 19]. A detailed discussion of parameter estimation in the MRCML models can be found in Adams *et al.* [12].

Modeling self-determination

We chose to illustrate the multidimensional analysis using the six dimensions of the TSRQ, as described above. It would also be possible to group the autonomous, controlled and amotivational items as three dimensions as is usually done in scoring the TSRQ, or the diet improvement and smoking cessation items as two separate dimensions, but we chose to examine these, as there were interesting patterns among the six dimensions. To enhance our illustration, we compare three approaches with the analysis of these data: the composite, consecutive and multidimensional approaches.

Composite approach

In the composite approach, the sum of scores received on the 30 items, ranging between 0 and 180, could be treated as the indicator for a single estimate of participants' perception of self-determination. The probability of a response in one of the seven categories, as opposed to the previous category, for each of the 30 TSRQ items is given by:

$$\log(P_{ik}/P_{ik-1}) = \theta - \delta_{ik}, \tag{2}$$

where θ and δ are as defined above.

Analyzing the data compositely with the software package ConQuest [17] produces estimates for a total of 181 parameters—180 item and step difficulties (one average item difficulty and five estimated steps from one response category to the next for each item; with seven categories, items are constrained to have only six identifiable parameters) and one population variance. In the analysis, we constrained the population mean to zero so that all item parameters could be estimated while ensuring parameter identification.

The composite approach has the advantage of parsimony in modeling the participant perspective, summarizing self-determination with a single number and its associated standard error. Furthermore, the reliability of participant self-determination estimates is quite high at 0.85 (calculated using the MML formulation [20]). Yet a clear disadvantage is that the differential information about self-determination relative to the six dimensions of reasons for healthy behavior is not represented in these results.

Consecutive approach

In the consecutive approach, the sum of scores on items associated with the six dimensions of reasons for healthy behavior are treated as six different indicators and modeled separately in different analyses. Separate models are used for each dimension using ConQuest. The difference between the composite and consecutive approaches is depicted conceptually in Fig. 3. In the consecutive approach, 186 parameters would be estimated, such that each dimension had one variance, one parameter for each item and one parameter for each estimated step. For example, for the ID dimension, we had one variance, six item parameters and 30 step parameters (six items × five steps). For INT, we had one variance, four item parameters and 20 step parameters. Likewise for INTR, IJ, EC and AMOT, there was one variance each, one parameter for each item (2 + 8 + 4 + 6 = 20) and one parameter for each step (20 × 5). As in the composite approach, the population means were all constrained to be zero.

The consecutive approach has the advantage of producing self-determination estimates and standard errors for all six dimensions of reasons for healthy behavior. Yet the consecutive approach ignores the possibility that self-determination across the six variables might be interrelated. Because the number of items defining each dimension is necessarily smaller, the standard errors of the consecutive estimates are larger than those resulting from a composite analysis. Reliabilities from the consecutive analyses range from 0.50 to 0.83 (see Table II), which are all smaller than the composite reliability of 0.85.

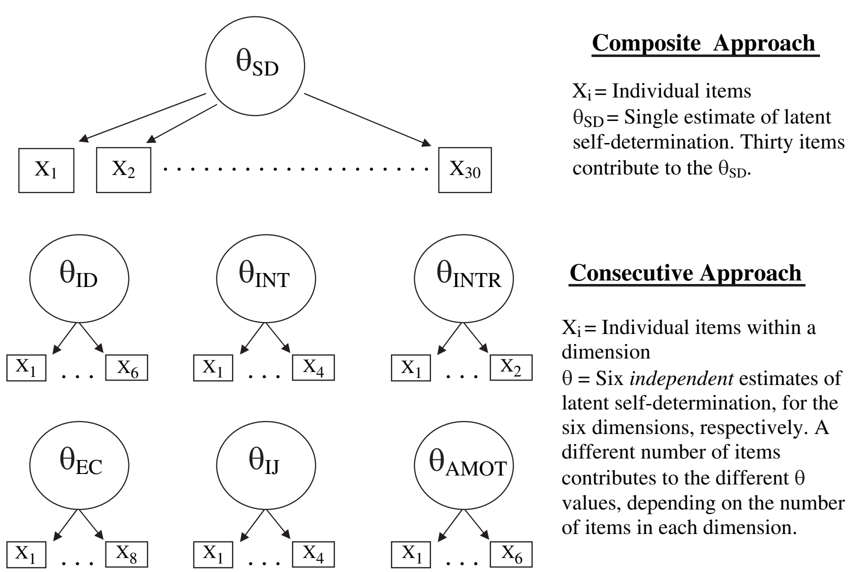

Fig. 3. Modeling self-determination: composite and consecutive approaches (after [14]).

Multidimensional approach

The multidimensional approach can be viewed as a compromise between the composite and consecutive approaches, one that incorporates the best of both approaches. The scores on each dimension provide distinct information about each participant, yet by incorporating the correlation between the dimensions directly into the model; the reliability for each of the six self-determination dimensions is closer to the composite reliability. Figure 4 illustrates the multidimensional approach. Note how there is a direct influence of the latent attitude for each dimension on the items assigned to that dimension depicted through the straight arrows, but that there is also an influence from all the other dimensions through the curved lines (which represent associations rather than causes). This approach can be modeled using the MRCML model to estimate latent attitudes across the six self-determination dimensions simultaneously. The multidimensional approach uses the formulation from Equation 1, repeated here for convenience (compare with Equation 2):

$$\log(P_{ik}/P_{ik-1}) = \theta_d - \delta_{ik}. \qquad (3)$$

Note that the major difference between Equations 2 and 3 is that here the person estimate, θ_d, is now also subscripted by the dimension, d. The multiple d values allow the researcher to model participants' level of attitude on each of the separate dimensions represented in the scale.

There are now six population means for this analysis (constrained to zero for identification), six variance estimates and 15 covariance estimates. Altogether, a total of 201 parameters are estimated using ConQuest.

Because the multidimensional approach is hierarchically related to the composite approach (i.e. the models are nested), the model fit can be

Table II. *Comparing reliabilities of three analyses of the self-determination data*

Dimension	Consecutive approach	Multidimensional approach	Composite approach
ID	0.77	0.83	NA
INT	0.75	0.84	NA
INTR	0.50	0.84	NA
EC	0.83	0.88	NA
IJ	0.78	0.85	NA
AMOT	0.60	0.70	NA
Self-determination (whole scale)	NA	NA	0.85

NA = not applicable.

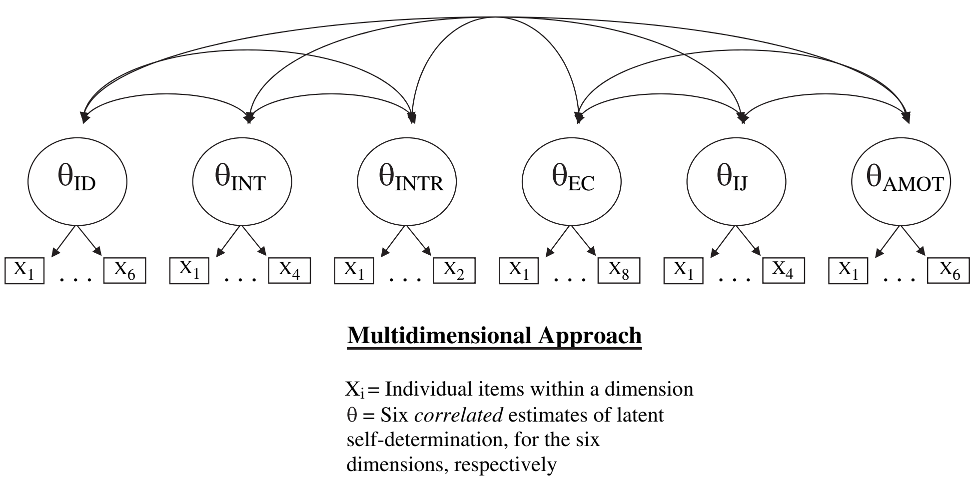

Fig. 4. Modeling self-determination: multidimensional approach, showing different numbers of items attributable to each dimension.

compared using the change in the deviance (G^2) value (the change in deviance indicates 'relative' model fit). The difference in deviance between the two nested models will approximate a χ^2 distribution with the number of degrees of freedom equal to the difference in the number of parameters estimated by the two models. Using the data from Table III, the difference in deviance between the two models is 6452.287, and the difference between the number of parameters is 20. This indicates that the multidimensional model fits the data significantly better than the composite model. On the basis of a comparison of Akaike's Information Criterion (AIC) [21], a method of comparing non-nested models analyzing the same data, the multidimensional model fits the data better than both the composite and the consecutive model (i.e. the AIC value is lower for the multidimensional approach). These indicators of statistical significance lead one to expect that there will also be differences between the models in terms of effect size (i.e. at the interpretational level). We illustrate several of these below with respect to (i) reliability, (ii) estimated correlations among dimensions and (iii) self-determination estimates under the different models.

Interpreting multidimensionality

Earlier the claim was made that an advantage of the multidimensional approach is an improvement in reliability relative to the consecutive approach. Table II supports this claim by showing the reliability for each of the six self-determination dimensions when using the two approaches. Under the multidimensional approach, the reliability for each self-determination dimension comes closer to the composite reliability of 0.85 (we use this reliability as the standard because it incorporates all the items possible in this scale). The most notable difference is for the dimension labeled INTR, which only had two items of its own (one from each of the contexts). Its reliability is considerably enhanced (error due to randomness of responses to items is diminished) by correlational information from responses to the other items in the multidimensional analysis. Components that had a larger number of items, such as EC with eight, had a smaller increase in reliability. Wang *et al.* [13] provide additional examples of reliability enhancement through multidimensional analysis. Note that these increases are conditional on using all the information available in the scale: if, say, the INTR results were to be recalculated using just the two INTR items, then the reliability would fall back to a level similar to that for the consecutive approach.

The multidimensional and consecutive approaches can also be compared with respect to estimated correlations between the self-determination dimensions (Table IV). ConQuest calculates both the correlation and covariance between the dimensions automatically in a multidimensional analysis. In this case, the correlation between the six dimensions ranged from –0.65 (between INTR and AMOT) to 0.05 (between EC and ID) to 0.98 (between INT and INTR). The highest positive correlations are between dimensions within the autonomous set, as might be expected by the

Table III. *Comparing three analyses of the self-determination data*

Approach	No. of parameters	G^2	AIC
Composite	181	73000.408	73362.408
Consecutive (six combined)	186	69107.112	69479.112
Multidimensional	201	66548.121	66950.121

Multidimensional compared with composite approach: likelihood ratio test (G^2): $\chi^2 = 6452.287$, df = 20, $p < 0.00001$.
Multidimensional compared with consecutive approach (AIC): 66950.121 < 69479.112.

Table IV. *Correlations between self-determination dimensions*

	ID	INT	INTR	EC	IJ	AMOT
ID	1.0	0.75	0.75	0.05	0.22	−0.37
INT	0.97	1.0	0.76	0.10	0.24	−0.41
INTR	0.97	0.98	1.0	0.11	0.25	−0.42
EC	0.05	0.14	0.13	1.0	0.63	0.13
IJ	0.27	0.32	0.33	0.82	1.0	0.01
AMOT	−0.58	−0.62	−0.65	0.19	0.00	1.0

Multidimensional correlations are shown below the diagonal; consecutive correlations above the diagonal.

relationship depicted in Fig. 1. Negative correlations are noted between the amotivated dimension and the three autonomous dimensions, a surprise, since the controlled dimensions, not the amotivated dimension, were hypothesized to lie at the opposite end of the scale from the autonomous dimensions. The controlled and autonomous sets of dimensions are relatively uncorrelated. So a person who chooses mostly 7s ('very true') on items associated with autonomous dimensions will generally choose lower numbers (not true) on items indicating amotivation, but will have no identifiable pattern of response on items associated with controlled dimensions. Whether or not these correlation statistics lead to revision of ideas about self-determination, they clearly indicate that modeling the data with a single composite dimension will not do justice to the interesting complexity of the scale, and hence, these results should be considered when designing mediator studies. When calculated using the consecutive approach, the correlations between the case estimates for all six dimensions are smaller (closer to zero) than when calculated using the multidimensional approach. Unlike the variance–covariance matrix produced by ConQuest using the multidimensional model, the consecutive approach gives correlations that are attenuated due to measurement error.

Using only the multidimensional estimates of participant attitudes about their self-determination, a comparison of two of the dimensions is plotted in Fig. 5. Both sets of θ values were standardized by creating z values [(case estimate – mean of case estimates)/standard deviation of case estimates] so that they could be compared directly. Note that the data are not continuous: each possible integer raw score value (0–13 for INTR and 0–24 for IJ) has a corresponding logit value. Positive values indicate that these participants chose higher response categories indicating the item statements were 'more true' for them. Participants' perception of self-determination along the autonomous dimension INTR and the controlled dimension IJ shows the low correlation, 0.33, between these two dimensions. Note that participants with a raw score of 12 out of 12 on INTR could have any of the whole

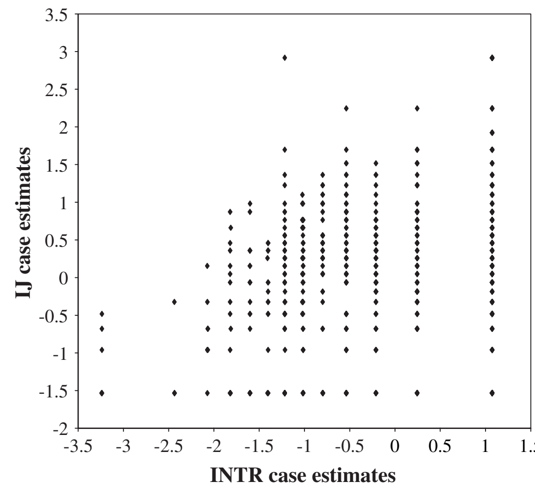

Fig. 5. Standardized estimates of participants' perception of self-determination on two dimensions.

range of raw scores from 0 to 24 on IJ. Which participants thus have greater self-determination? Multidimensional analysis has allowed us a closer look at the data, with two possible interpretive consequences. First, the items in each of these dimensions may require additional examination to see if they indeed represent autonomous and controlled reasons for healthy behavior as the authors of the TSRQ projected [6]; if not, modification of the items might result. Second, the idea that autonomous and controlled reasons fit on opposite ends of a continuum in a construct for self-determination [5] may not hold for all these dimensions; refinement of the construct might result. In any case, these empirical data allow a closer look at the interpretation possible for the common practice of averaging over autonomous and controlled items of the TSRQ and finding the difference between them.

Another way to compare the self-determination estimates across dimensions, d, is to use a sum of squares indicator, DI, for each participant, p [14].

$$DI_p = \sum_{d=1}^{6} (\bar{\theta} - \theta_d)^2$$

If we (somewhat arbitrarily) set the threshold for a discrepant case at DI_p = 1.0, there are 831 participants (89%) in this sample for whom dimensional estimates would reveal differing stories about participant perception of self-determination.

Analyzing examples of discrepant cases is an indication that there might be important diagnostic or interpretive ramifications of ignoring multidimensionality when assessing participant perception. For most participants in our sample, the dimensional scores look very different from a composite score. Consider, for example, two of the cases in our data, each having a raw score of 94 (out of 180 possible). Table V shows the differences in the raw scores across the dimensions that are obscured by reporting only the composite score. Participant number 522's responses showed more autonomous types of self-determination, compared with number 598's more controlled types of self-determination. The fact that the TSRQ is not generally reported as a single composite score is consistent with the problem noted here of comparing two participants with the same score; other scales may have similarly important interpretational differences between the composite and multidimensional results. The importance for the TSRQ is that, according to SDT, those who show more autonomous types of self-determination should have a better chance of changing behavior than anyone else [5, 6]. Reliance only on some composite estimate may falsely identify participants that will successfully change behavior. Comparing the multidimensional estimates with evidence of behavior change may support the theory and can help guide appropriate intervention. Note that in the multidimensional analysis, participant 598's INTR dimension is strongly influenced by the other dimensions, which should be expected because it has the fewest items. The ID dimension is also somewhat influenced by other dimensions in the multidimensional analysis of both cases.

Discussion

In much of the behavioral sciences, there is a desire to either (i) disaggregate the scores from a scale into subscales and report them as separate dimensions of participant attitude or perception or (ii) to aggregate the scores of subscales and report this as a single dimension of perception or outcome. Both these scenarios depart from the measurement ideal of single-dimension scales. The MRCML models are useful tools when the objective is to measure more than one latent construct. In this paper, we analyzed the TSRQ as a single dimension to illustrate the composite approach, as separate and isolated dimensions to illustrate the consecutive approach and as a scale with multiple correlated dimensions to illustrate the multidimensional approach using a MRCML model. Our example taken from the BCC data on the TSRQ highlights some important differences between the three approaches. The essential difference is that the consecutive approach is simply a single-dimensional model repeated a number of times using subsets of the full range of items on a given scale. Because there are fewer items, just those items included that define each latent domain, the standard error of measurement

Table V. *Comparing ability estimates in logits (standard error) of two participants with the same self-determination score of 94/180*

	θ_{ID}	θ_{INT}	θ_{INTR}	θ_{EC}	θ_{IJ}	θ_{AMOT}	$\theta_{SD\text{-composite}}$
Participant 522							
Raw score (% of total)	19	12	7	4	5	5	52
Consecutive estimate	−0.27 (0.58)	0.21 (0.51)	0.35 (0.81)	0.00 (0.26)	0.24 (0.29)	0.38 (0.22)	0.14 (0.11)
Multidimensional estimate	−0.44 (0.57)	0.12 (0.52)	0.30 (0.97)	−0.03 (0.26)	0.22 (0.30)	0.38 (0.22)	
Participant 598							
Raw score (% of total)	18	7	4	9	9	4	52
Consecutive estimate	−0.54 (0.49)	−0.91 (0.30)	−0.71 (0.38)	0.45 (0.21)	0.79 (0.31)	0.33 (0.22)	0.14 (0.11)
Multidimensional estimate	−0.71 (0.49)	−1.02 (0.30)	−1.10 (0.41)	0.43 (0.22)	0.82 (0.32)	0.34 (0.22)	

for person estimates increases and the reliability of estimates does not attain that of the full composite model. In other words, the multidimensional approach results in less error than the consecutive approach. In addition, the lower deviance scores with the multidimensional approach indicate less error than the composite approach when comparing model fit. Thus, we present the multidimensional approach as an improvement over both the composite and consecutive approaches. The approach provides distinct estimates for multiple latent constructs, yet by modeling the dimensions as interrelated, the reliability of the estimates comes closer to that found under the much less informative composite analysis.

Structural Equation Modeling (SEM) is another method that allows for multidimensional analysis of scales. Apart from the aim to be consistent with the approach used in earlier articles in this special issue [3, 4], item response modeling differs from SEM in that (i) it models the actual response data rather than the covariance matrix among the variables, and hence, (ii) it allows a finer grain of interpretation and fit analysis. Among item response models, the MRCML model was chosen because, as a direct extension of the same models used to analyze a scale via the composite and consecutive approaches, the MRCML provides a consistency to the argument for similarities and differences between the three approaches.

Beyond the statistical rationale for a multidimensional approach, we believe that there are important interpretational differences as well, particularly for the behavioral sciences. Treating participant-reported data that are multidimensional in nature as if they represent a single dimension may misrepresent participant's perceptions about their self-determination, for example. This is less of a danger when the dimensions in question have a moderately high positive correlation, as in the case of the autonomous motivation dimensions in our example; one dimension could represent all three. Nonetheless, we show that participants can have their perceptions or outcomes misrepresented, particularly if the scale includes dimensions that have small or negative correlations.

The data example used in this paper is a fairly simple approach to modeling both unidimensional and multidimensional self-determination. We ignore what is a possible violation of local item independence in the 'bundling' of items within a common prompt (i.e. because we ask the same exact questions for both diet and smoking). This analysis is not an attempt to draw inferences in any absolute sense about participant-reported self-determination or the use of the TSRQ. Rather, it has been meant as an illustration of the 'relative' merits of a multidimensional approach to a composite or consecutive approach in the context of the behavioral sciences.

Acknowledgements

We would like to thank Louise C. Mâsse formerly of the National Cancer Institute (NCI), now from the University of British Columbia, for organizing the project on which this work is based and for providing crucial guidance throughout the writing of the paper. We thank Geoff Williams for directing us in interpreting the scale and its intent. Support for this project was provided by NCI (Contract No. 263-MQ-31958). We thank the BCC for providing the data that were used in the analyses. Any errors or omissions are, of course, solely the responsibility of the authors.

References

1. Williams GC, Grow VM, Freedman ZR *et al*. Motivational predictors of weight loss and weight-loss maintenance. *J Pers Soc Psychol* 1996; **70**: 115–26.
2. Ory MG, Jordan PJ, Bazzarre T. The Behavior Change Consortium: setting the stage for a new century of health behavior-change research. *Health Educ Res* 2002; **17**: 500–11.
3. Wilson M, Allen DD, Li JC. Improving measurement in health education and health behavior research using item response modeling: introducing item response modeling. *Health Educ Res* 2006; **21**(Suppl 1): i4–i18.
4. Wilson M, Allen DD, Li JC. Improving measurement in health education and health behavior research using item response modeling: comparison with the classical test theory approach. *Health Educ Res* 2006; **21**(Suppl 1): i19–i32.
5. Williams GC, Minicucci DS, Kouides RW *et al*. Self-determination, smoking, diet and health. *Health Educ Res* 2002; **17**: 512–21.

6. Williams GC, Deci EL. Internalization of biopsychosocial values by medical students: a test of self-determination theory. *J Pers Soc Psychol* 1996; **70**: 767–79.
7. Reckase MD. The difficulty of test items that measure more than one ability. *Appl Psychol Meas* 1985; **9**: 401–12.
8. Reckase MD, McKinley RL. The discriminating power of items that measure more than one dimension. *Appl Psychol Meas* 1991; **15**: 361–73.
9. Ackerman T. A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *J Educ Meas* 1992; **29**: 67–91.
10. Ackerman T. Using multidimensional item response theory to understand what items and tests are measuring. *Appl Meas Educ* 1994; **7**: 255–78.
11. Embretson SE. A multidimensional latent trait model for measuring learning and change. *Psychometrika* 1991; **56**: 495–515.
12. Adams RJ, Wilson M, Wang W. The multidimensional random coefficients multinomial logit model. *Appl Psychol Meas* 1997; **21**: 1–23.
13. Wang W-c, Wilson M, Adams RJ. Rasch models for multidimensionality between items and within items. In: Wilson M, Engelhard G, (eds). *Objective Measurement: Theory into Practice*, vol. IV. Norwood, NJ: Ablex, 1997, 139–55.
14. Briggs DC, Wilson M. An introduction to multidimensional measurement using Rasch models. *J Appl Meas* 2003; **4**: 87–100.
15. Baranowski T, Allen DD, Mâsse L *et al.* Does participation in an intervention affect responses on self-report questionnaires? *Health Educ Res* 2006; **21**(Suppl 1): i98–i109.
16. Wright BD, Masters GN. *Rating Scale Analysis.* Chicago, IL: MESA Press, 1982.
17. Wu ML, Adams RJ, Wilson MR. *ACER ConQuest: Generalised Item Response Modelling Software* [computer program]. Hawthorn, Australia: ACER (Australian Council for Educational Research) Press, 1998.
18. Dempster AP, Laird NM, Rubin DB. Maximum likelihood estimation from incomplete data via the EM algorithm. *J R Stat Soc* 1977; **39**: 1–38.
19. Bock RD, Aitken M. Marginal maximum likelihood estimation of item parameters: an application of the EM algorithm. *Psychometrika* 1981; **46**: 443–59.
20. Mislevy RJ, Beaton AE, Kaplan B *et al.* Estimating population characteristics from sparse matrix samples of item responses. *J Educ Meas* 1992; **29**: 133–61.
21. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974; **19**: 716–23.

Received on October 20, 2005; accepted on July 19, 2006