Evaluation of Social Cognitive Scaling Response Options in the Physical Activity Domain

Ryan E. Rhodes, Deborah Hunt Matheson, & Rachel Mark

2010

Running Head: Social Cognition Scaling

Evaluation of Social Cognitive Scaling Response Options in the Physical Activity Domain

Abstract

The purpose of this study was to compare the reliability, variability, and predictive validity of two common scaling response formats (semantic differential, Likert-type) and two numbers of response options (five-point, seven-point) in the physical activity domain. Constructs of the theory of planned behavior (TPB) were chosen in this analysis based on its high frequency of application in exercise and physical activity. Participants were 412 undergraduate students who completed measures of the TPB and self-reported physical activity two weeks later. One of four questionnaires, each containing a scaling response format, were distributed randomly and formed four groups of approximately n= 100 for comparisons (five-point Likert, five-point semantic differential, seven-point Likert, seven-point semantic differential). Results showed that the seven-point options had greater variability than the five-point options and that the seven-point Likert scale had the highest overall reliability. These differences, however, did not translate into predictive validity of behavior. The findings support the use of all these types of scales with physical activity research because of their relatively equivalent outcomes.

Key Words: Likert Scaling, Semantic Differential Scaling, Theory of Planned Behavior, Physical Activity

The physical and psychological health benefits of regular physical activity are convincing and well substantiated (Warburton, Katzmarzyk, Rhodes, & Shephard, 2007); yet most of the populace does not engage in enough activity to reap these benefits (Canadian Fitness and Lifestyle Research Institute, 2002; U.S. Department of Health and Human Services, 2003). As a result, the promotion of regular physical activity is a public health priority. In order to change physical activity behavior patterns it has been suggested that a sound understanding of its determinants, or mediators, is required in order to target and hone effective interventions (Baranowski, Anderson, & Carmack, 1998; Lewis, Marcus, Pate, & Dunn, 2002). To this end, the testing of various theories has been prolific in an attempt to explain behavior and behavior change (Noar & Zimmerman, 2005; Rhodes & Pfaeffli, in press).

The general goal of this field of research is to establish robust and substantive theoretical constructs in explaining physical activity behavior. It is generally predicated on demonstrating large R^2 values with behavior (Baranowski et al., 1998; Noar & Zimmerman, 2005) using multiple regression equations with a series of predictors. Thus, the focus is at the structural level of the prediction model. Constructs are compared for independent predictive validity based on theoretical hypotheses.

Less work has been performed at the measurement-level of these models, particularly the scaling used for these constructs. Although perhaps less glamorous than the structural-level validity testing, measurement and scaling is considered foundational to the process. The measurement response options give clues to the participants in terms of how and where to answer (Menon, Raghubir, & Schwarz, 1995; Schwarz, 1999; Schwarz, Hippler, Deutsch, & Strack 1985) while the number of response options creates the variance of the measure (Aiken &

Groth-Marnat, 2006; DeVellis, 1991). Both of these factors are critical, in conception, to the validity of the construct being measured (Messick, 1995).

Evidence is mixed, however, as to the impact that scaling has on the variance of constructs when comparing reasonable methods and response options (Aiken & Groth-Marnat, 2006). Extreme scaling options certainly affect variance and response judgements (Schwarz, 1999), and this has been replicated in the physical activity domain when measuring behavior (Courneya, Jones, Rhodes, & Blanchard, 2003). Furthermore, open-ended scaling generally results in superior variance and better predictive validity compared to fixed scaling approaches (Menon et al., 1995) and this has also been replicated in the physical activity domain (Rhodes, Matheson, & Blanchard, 2006). The more subtle aspects of scaling, however, have seen less evidence for marked differences.

In terms of response options, Lam and colleagues (Lam & Klockars, 1982; Lam & Stevens, 1994) have demonstrated limited differences in variability when full-range (i.e., equal response options in negative and positive evaluation) or "packed" range (more positive or negative response options) response options are provided and this has similar support in the exercise domain (Courneya, Conner, & Rhodes, 2006). The notion here is that people can adjust to the psychological width of the options provided and it has limited impact on total variance. In terms of the total number of response options, the issue has been one of debate since very early scaling efforts. Likert (1932) and Remmers and Ewart (1941), for example, found that reliability and variability increase as a function of additional response options, while others have not replicated this finding (Bendig, 1951; Mattell & Jacoby, 1972). Still, measurement pundits have found that limited variability in measurement is gained beyond providing seven response options

(Aiken & Groth-Marnat, 2006; Johnson & Dixon, 1984) and this has also been supported in exercise research (Courneya et al., 2006).

Nevertheless, a comparison of more typical scaling practices in exercise psychology research has received scant attention and it is important to tease-out any measurement artefacts that may account for different variance in contemporary measures. Courneya et al. (2006) (7 to 11 point scales) and Rhodes, Matheson and Blanchard (2006) (open scaled measures) manufactured their scaling for the purpose of those specific studies; neither approach represents common scaling used in prediction research. Our review of the most prominent models applied to physical activity, namely social cognitive theory (Bandura, 1998), theory of planned behavior (Ajzen & Driver, 1991), and the transtheoretical model (Prochaska & DiClemente, 1982) showed that two scaling styles and two response widths are most common in the extant literature[1]. First, for scaling style, response options tended to be operationalized as semantic differential type phrasing (e.g., extremely unconfident to extremely confident) or via Likert type scaling (e.g., strongly disagree to strongly agree) based on an item stem with polarized content. Second, for number of response options, most assessments used either five-point or seven-point options.

Establishing the differences in variability and predictive validity with behavior are present when using these common options is important. Researchers focused on scaling have highlighted the importance of the measurement topic/domain on the outcome (Frisbie & Brandenburg, 1979; Lam & Stevens, 1994). Thus, tests in the physical activity domain are necessary and even specific constructs may vary by scaling option. Second, the notion of scaling choice is a common commentary among manuscript and grant reviewers; many often having strident opinions but very limited research to draw upon. An evaluation of common scaling styles would help shed light on this sometimes contentious issue. Finally, as variability and

predictive validity is often the cornerstone in tests of mediation and physical activity prediction, it stands to reason that a solid understanding of how scaling style affects these factors is warranted.

Therefore, the purpose of this study was to compare two common response style options (semantic differential, Likert) and two common numbers of response options (five-point, seven-point) for variability, reliability and predictive validity with physical activity behavior. The model chosen for these analyses was Ajzen's (1991) theory of planned behavior (TPB) based on its prolific use and validation in the physical activity domain (Symons Downs & Hausenblas, 2005) and its general use of all four scaling options in the extant literature. Briefly, the TPB proposes that behavior is a result of one's intention to act (i.e., motivation, readiness) and perception of behavioral control (i.e., ability to perform the behavior when motivation is held constant). Intention is subsequently a function of one's affective (e.g., enjoyability evaluation) and instrumental (e.g., utility evaluation) attitudes about the behavior, subjective norm (i.e., perceived pressure to perform the behavior from others) and perceived behavioral control.

Based on prior literature (Likert, 1932; Remmers & Ewart, 1941), it was hypothesized that seven-point scaling may demonstrate better variability and reliability compared to the five-point options if any differences were to emerge. The researchers were unable to find research that had evaluated semantic differential and Likert options in any domain, but it seemed these options should be relatively equivalent in predictive validity of intention and behavior based on work by Lam and Stevens (1994). Thus, the null hypothesis was considered for these scaling option formats.

Method

Participants and Procedure

Five hundred and eighteen students volunteered to participate in the study during their introductory psychology (classes n = 20) or physical education (classes n = 7) courses. Informed consent was obtained from the participants and the study received IRB approval. Participants attended large group sessions during September and October 2007 and 2008, completing self-report measures of the TPB and past physical activity. Participants received one of four possible questionnaires that were randomly distributed in the sessions; this resulted in four independent groups of approximately n = 120. These questionnaires differed by the scaling used for the TPB. The possible stems were: 1) five-point Likert (strongly disagree, disagree, undecided, agree, strongly agree) 2) five-point semantic differential (positive/negative adjective with extremely, the adjective and undecided), 3) seven-point Likert-type (strongly disagree, moderately disagree, slightly disagree, undecided, slightly agree, moderately agree, strongly agree), and 4) seven-point semantic differential (positive/negative adjective with extremely, moderately, slightly, and undecided). Participants were then asked to complete a behavioral measure two-weeks after the initial questionnaire administration. All participants completed this follow-up in class.

Instruments

The definition of physical activity (PA) was chosen to reflect Health Canada's position stand for recommended weekly activity among adults (Health Canada, 2002). PA was defined as activities performed with at least moderate intensity, 4 or more times per week, accumulating at least 30 minutes each time. Participants were asked to consider this definition when answering the TPB items.

Attitude toward regular physical activity. Three items were used to tap the instrumental (e.g., useful-useless, wise-unwise, beneficial-harmful) and three items were used to tap the affective (enjoyable-unenjoyable, pleasant-unpleasant, exciting-boring) aspect of attitude as

suggested by Ajzen (2002). The statement that preceded the adjectives was “For me, regular physical activity over the next 2 weeks would be.” Likert-style and semantic differential style scaling; however, differed in how this was approached. For Likert-style scaling, the positive valence adjectives were used (e.g., enjoyable) and immediately followed by strongly disagree to strongly agree response options. For the semantic differential scaling the adjectives were embedded within the response options as defined in this approach (see Table 3 for reliabilities).

Subjective norm was measured by four items; two items measured the injunctive component of subjective norm and two items assessed the descriptive component of subjective norm based on the recommendation of Ajzen (2002). These components were aggregated to form a scale based on the findings of Rhodes and colleagues (2006; 2003a). The items were: (1) “Most people who are important to me would encourage me to engage in regular physical activity over the next 2 weeks," 2) “Most people whose opinions I value would approve of me engaging in regular physical activity over the next 2 weeks,” 3) “Most people who are important to me will be…” and 4) “I think that over the next 2 weeks, the activity levels of most people who are important to me will be...” Semantic differential scaling used discourage-encourage, disapprove-approve, inactive-active, and low-high for items one through four respectively. By contrast, the Likert-type responses included “active” and “high” on the end of item stems three and four and then provided strongly disagree to strongly agree response options (see Table 3 for reliabilities).

Perceived behavioral control was measured by three questions recommended by Rhodes and Courneya (2003b; 2004). The three PBC items were: 1) “In the next 2 weeks, doing physical activity, if I really wanted to, is under my control”; 2) “I am confident I could engage in regular physical activity in the next 2 weeks if I wanted to”; and 3) “Engaging in regular physical activity over the next 2 weeks if I wanted to do so would be easy…” Semantic differential items

included uncontrollable-controllable, unconfident-confident, and difficult-easy response options for items one through three, respectively while responses for the Likert-type scaling provided strongly disagree to strongly agree options (see Table 3 for reliabilities).

Intention was measured using items from Rhodes, Blanchard, Matheson and Coble (2006) that focused on the motivational domain conceptualized by Ajzen (1991). These items included: 1) "I am committed to engage in physical activity over the next 2 weeks" 2) "I am motivated to engage in regular physical activity over the next 2 weeks," and 3) "I am determined to engage in regular physical activity over the next two weeks." Semantic differential items included uncommitted-committed, unmotivated-motivated, and undetermined-determined response options for items one through three respectively while responses for the Likert-type scaling provided strongly disagree to strongly agree options (see Table 3 for reliabilities).

Physical activity behavior was measured using the Godin Leisure Time Exercise Questionnaire (Godin, Jobin, & Bouillon, 1986; Godin & Shephard, 1985). The instrument contains three open ended questions covering the frequency of mild (e.g., easy walking), moderate (e.g., fast walking), and strenuous (e.g., jogging) exercise completed during free time for at least 30 minutes duration in a typical week and it has been validated favorably with accelerometry, fitness levels, and other self-report measures (Jacobs, Ainsworth, Hartman, & Leon, 1993). Our adaptation included substituting "physical activity" for "exercise" and over the last two weeks instead of one week. Strenuous and moderate physical activity frequencies were aggregated to produce a total activity frequency at or above moderate intensity. Mild activity was not included as an indicator due to its incongruence with our definition of regular physical activity.

Analysis Plan

Initial tests of randomization included analysis of variance tests of past and future physical activity, age, and year in university across groups and Chi-square tests of group by gender. Differences in reliability (Chronbach's alpha), TPB construct correlations with intention and physical activity behavior, and the multiple R for the TPB model to test predictive validity across groups were conducted using the test for independent associations (Hedges & Olkin, 1985) recommended by Glass and Hopkins (1996). The multiple R was calculated for the prediction of intention and behavior with ordinary least squares regression based on the proposed structure of the theory of planned behavior (Ajzen, 1991). Specifically, intention and perceived behavioral control were used as independent variables to predict behavior, while affective and instrumental attitude, subjective norm, and perceived behavioral control were subsequently used as independent variables to predict intention. Significant differences in correlations were followed-up using Fisher *Z* tests for assessment of bivariate associations similar to using Tukey HSD in analysis of variance procedures. Difference in variance across groups was assessed via Levene's test of equal variances. Type one error was set at .05 throughout and the analyses were powered to detect a small effect size ($f = .17$, $q = .19$).

Results

A total of 412 students completed the follow-up measure and were subsequently used in the analyses. Ninety-two participants did not return for the follow-up administration and 14 participants could not be matched with a corresponding follow-up measure (code incorrect or indecipherable). There were no differences between drop-outs and participants on baseline physical activity, any of the TPB variables, demographics (age, gender) or group assignment ($p > .05$). Thus, these data can be considered missing completely at random and not problematic to sample generalization (Allison, 2002). Normality assumptions were met with these data. Mean

age of the sample was 22.17 (*SD* = 6.69), 67% were female, the mean year in college was 1.87 (*SD* = 1.20), and 70% were meeting Canada's physical activity guidelines (Health Canada, 2002) at baseline. Randomization of the four groups was also successful with 111 in the five-point Likert condition, 102 in the five-point semantic differential condition, 99 in the seven-point Likert-type condition and 100 in the seven-point semantic differential condition; no differences were observed across age, gender year in school and past or two-week physical activity by condition (see Table 1).

Means and variances of the Theory of Planned Behavior (TPB) constructs across the scale response options are presented in Table 2. All TPB constructs showed differences in variance across conditions ($p < .01$). Subjective norm, perceived behavioural control (PBC), and intention all showed that the five-point scale options had lower variance than the seven-point response options. For affective attitude, however, five-point semantic differential had a smaller variance compared to all groups while five-point Likert was smaller than both seven-point options. Instrumental attitude was similar, but in this case, the five-point Likert had a smaller variance compared to all other conditions and five-point semantic differential was only different from seven-point Likert-type.

Examination of the reliabilities across scale response options can be found in Table 3. With the exception of subjective norm, all other TPB constructs were significantly different in reliability ($p < .01$). Each construct, however, had a slightly different profile of differences in the post-hoc assessments. Affective attitude showed lower reliabilities for the semantic differential scale options compared to Likert-type options. Instrumental attitude had a lower reliability for five-point Likert compared to five-point semantic differential and seven point-Likert, while seven–point Likert-type was also more reliable than its comparable seven-point semantic

differential scale option. Finally, seven-point Likert-type was more reliable than seven-point semantic differential and five-point Likert for PBC and a similar finding was identified for intention but seven-point Likert was also more reliable than five-point semantic differential in this case.

Comparisons for the relationships between TPB constructs and intention and physical activity behavior by condition can be found in Table 4 and 5, respectively. For intention relationships by scale response option, only affective attitude was significant ($p < .01$). Post hoc tests identified that the seven-point Likert-type response option was significantly larger than all other options ($p < .05$). This extended to the multivariate equation with the seven-point response option showing a significantly larger effect compared to both semantic differential options. Of key importance; however, is that this pattern was not observed with tests of relationships with behavior. Indeed, there were no significant differences by scaling response options between the TPB and behavior (see Table 5). Finally, TPB models predicting intention and behavior are detailed in Table 6 and Table 7 respectively. Affective attitude was the significant predictor of intention with varying smaller contributions from other constructs (affective attitude $\beta = .42$ to .75; instrumental attitude $\beta = -.12$ to .20, subjective norm $\beta = .13$ to .25; PBC $\beta = .08$ to .21). Intention was the significant predictor of behavior in the four conditions within the multivariate model (intention $\beta = .51$ to .57; PBC $\beta = -.03$ to .18). There was some evidence of multicollinearity in the form of negative coefficients in the 7-point scaling options although these coefficients were not significant and tolerance levels for these models were within acceptable limits (i.e., > .20) (Field, 2005).

Discussion

The purpose of this study was to compare the reliability, variability, and predictive validity of two common scaling response formats (semantic differential, Likert-type) and two numbers of response options (five-point, seven-point) in the physical activity domain. Constructs of the theory of planned behavior were chosen in this analysis based on its high frequency of application in exercise and physical activity (Symons Downs & Hausenblas, 2005). Research on scaling response options is scant in the physical activity domain, but the researchers hypothesized that additional variability and subsequently higher reliability may be present in the seven-point scaling options. The null hypothesis was considered in regard to an effect between response formats and the predictive effect on behavior although the exploratory nature of this study was acknowledged. Overall, the findings were interesting and generally supported these hypotheses.

First, the hypothesis that the seven-point scaling options would demonstrate larger variability compared to the five-point scaling options was supported. All TPB constructs exhibited evidence for larger variance in the seven-point scaling options. This makes theoretical sense because the seven-point options provide more fidelity to grade one's answer (Likert, 1932). While this has not been found in all studies on measurement response options (Bendig, 1951; Mattell & Jacoby, 1972), it does match prior work comparing seven and 11 point response options in the physical activity domain (Courneya et al., 2006). Thus, seven point scales in the physical activity domain offer more variability than five-point scale options.

The hypothesis that this would translate to better reliability (i.e., internal consistency), however, only had partial support. In this case, it was only the seven-point Likert-type scales that showed relatively consistent reliability differences and these were often higher than the seven-point semantic differential scaling. This is the first study to examine differences between Likert-

style response options and semantic differential options so the finding is interesting, but will benefit from replication. The explanation for this finding is that a mix of increased variability and common methods variance may be accounting for this difference. The Likert-type scaling maintains exactly the same response options throughout the measure administration while semantic differential scaling makes the participant consider a range of adjectives across the administration. This shift in scaling properties may lower the reliability because each item makes participants attend to responding on a slightly different metric.

Clearly the main outcome of interest for physical activity researchers is whether the additional variability and reliability noted above contributed to improved predictive validity. Models such as the TPB are tested by assessing explained variance and thus any additional predictive validity offered purely via scaling is important to understand. Further, these models are important considerations in intervention efforts (Baranowski et al., 1998; Rhodes & Pfaeffli, in press) for the same basic premise. The present results supported the null hypothesis. Indeed, scaling response options did not have an effect on the relationship of any TPB construct and its relationship with physical activity. In the prediction of intention, the affective attitude construct had a larger relationship in the seven-point Likert option compared to all other options and this translated to the multivariate prediction model; however, the effect was not observed with behavior supporting the common methods rationale given above in the discussion of reliability. Overall, the researchers conclude that the differences observed from increased reliability and variance did not translate to meaningful differences in the model.

From a practical point of view, this outcome should be viewed as a positive. The outcome suggests that the various scaling practices used in prior work are probably having a minimal effect on outcomes. The outcome also suggests that scaling response options can be tailored to

serve various methods when necessary and researchers can increase their confidence that the overall deviations will be negligible. For example, phone-based surveys might be easier to conduct with a five-point Likert option because it creates less memory and cognitive burden than semantic differential or seven-point options. Up until now, the effect the scaling format itself may have been having on the predictive validity of physical activity was speculative.

From a theoretical perspective, the evidence demonstrated that increases in reliability and variance do not always translate into meaningful differences in prediction models. False precision or spurious variance is always a possibility and seems to be the case in this study. Social cognitive theories have been criticized for factors like common methods variance (Ogden, 2003); therefore, studying differences in behavioural relationships on a different metric is important.

Future research directions for an evaluation of scaling in the physical activity domain may benefit by an evaluation of the interaction between scaling response options and the item phrases. While this has not been performed in physical activity, there is evidence that scaling response options and item phrases interact and that the interaction is domain specific (Lam & Stevens, 1994). A test of semantic differential scaling with only the end anchors compared to the scale where all points are described may be prudent. The scaling metric used also needs more attention. The reason of no differences in predictive validity may be all of these response options utilize the same genre of fixed statement followed by a graded metric. This type of scaling is referred to as quasi-interval level scaling (Glass & Hopkins, 1996) because the transparent width between response options is noticeably better than ordinal scaling but not pure interval. By comparison, gradations by percent effort or frequency utilize a ratio-level scale and may be able to improve meaningful precision of estimations. Initial research has supported this conclusion

with continuous open measures based on frequency in the physical activity domain (Rhodes, Matheson, et al., 2006) but the application has seen limited use.

The limitations of this study should be mentioned. First, the study utilizes an undergraduate sample comprised primarily of females. While age and gender are not moderators of physical activity and theory of planned behavior relations (Rhodes, Blanchard, & Blacklock, 2008), there may be some unforeseen contextual factors that limit sample generalizability to the populace. Second, the behavior measure administered was self-report and subject to biases though it is still an issue of debate how severe these are (Prince et al., 2008); factors such as truthfulness and social desirability have not been related to self-reported physical activity (Motl, McAuley, & DiStefano, 2005). Finally, this study employed a test of theory of planned behavior and the results may not generalize to other constructs in the physical activity domain. The researchers hopee that this theory can act as a reasonable proxy model for other related constructs, but differences in the measurement domains and contextual factors may limit complete generalizability.

In summary, reliability, variability, and predictive validity (of behavior) of two common scaling response formats (semantic differential, Likert-type) and two numbers of response options (five-point, seven-point) were compared in the physical activity domain. Results were that the seven-point options had greater variability than the five-point options and that the seven-point Likert scale had the highest overall reliability. These differences, however, did not translate into predictive validity of behavior. The findings support the use of all these types of scales with physical activity research.

1. Footnote: Measurement of self-efficacy is sometimes measured in numerical gradation (Bandura, 1997). This style was not assessed.

Acknowledgement

RER is supported by a New Investigator award from the Canadian Institutes of Health Research and with funds from the Social Science and Humanities Research Council of Canada and the Canadian Diabetes Association. DHM is supported by a Vancouver Island University Research Award. We also wish to acknowledge the hard work of data collection and data entry by Megan Yim.

References

Aiken, L. B., & Groth-Marnat, G. (2006). *Test design and construction* (12th ed.). Boston: Pearson Education.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211.

Ajzen, I. (2002). Constructing a TPB questionnaire: Conceptual and methodological considerations. Retrieved April 7, 2007, from http://www-unix.oit.umass.edu/~aizen/pdf/tpb.measurement.pdf

Ajzen, I., & Driver, B. L. (1991). Prediction of leisure participation from behavioral, normative, and control beliefs: An application of the theory of planned behavior. *Leisure Sciences, 13*, 185-204.

Allison, P. D. (2002). *Missing Data.* Thousand Oaks: Sage Publications.

Bandura, A. (1997). *Self-efficacy, the exercise of control*. New York: Freeman.

Bandura, A. (1998). Health promotion from the perspective of social cognitive theory. *Psychology and Health, 13*, 623-649.

Baranowski, T., Anderson, C., & Carmack, C. (1998). Mediating variable framework in physical activity interventions: How are we doing? How might we do better? *American Journal of Preventive Medicine, 15*, 266-297.

Bendig, A. W. (1951). Reliability and the number of rating scale categories. *Journal of Applied Psychology, 38*, 38-40.

Canadian Fitness and Lifestyle Research Institute. (2002). 2002 Physical Activity Monitor. Retrieved August, 2004, from http://www.cflri.ca/cflri/pa/surveys/2002survey/2002survey.html

Courneya, K. S., Conner, M., & Rhodes, R. E. (2006). Effects of different measurement scales on the variability and predictive validity of the "two-component" model of the theory of planned behavior in the exercise domain. *Psychology and Health, 21*, 557-570.

Courneya, K. S., Jones, L. W., Rhodes, R. E., & Blanchard, C. M. (2003). Effect of response scales on self-reported exercise frequency. *American Journal of Health Behavior, 27*, 613-622.

DeVellis, R. F. (1991). *Scale development: Theory and practice* (Vol. 26). London: Sage.

Field, A. (2005). *Discovering Statistics using SPSS* (2nd ed.). London, U.K: Sage.

Frisbie, D. A., & Brandenburg, D. C. (1979). Equivalence of questionnaire items with varying response formats. *Journal of Educational Measurement, 16*, 43-48.

Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Godin, G., Jobin, J., & Bouillon, J. (1986). Assessment of leisure time exercise behavior by self-report: A concurrent validity study. *Canadian Journal of Public Health, 77*, 359-361.

Godin, G., & Shephard, R. J. (1985). A simple method to assess exercise behavior in the community. *Canadian Journal of Applied Sport Science, 10*, 141-146.

Health Canada. (2002). Health Canada's Physical Activity Guide. Retrieved August, 2004, from http://www.hc-sc.gc.ca/english/lifestyles/physical_activity.html

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Jacobs, D. R., Ainsworth, B. E., Hartman, T. J., & Leon, A. S. (1993). A simultaneous evaluation of ten commonly used physical activity questionnaires. *Medicine and Science in Sports and Exercise, 25*, 81-91.

Johnson, W. L., & Dixon, P. N. (1984). Response alternatives in likert scaling. *Educational and Psychological Measurement, 44*, 563-566.

Lam, T. C. M., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement, 19*, 317-322.

Lam, T. C. M., & Stevens, J. J. (1994). Effects of content polarization, item wording, and rating scale width on rating response. *Applied Measurement in Education, 7*, 141-158.

Lewis, B. A., Marcus, B., Pate, R. R., & Dunn, A. L. (2002). Psychosocial mediators of physical activity behavior among adults and children. *American Journal of Preventive Medicine, 23*((2S)), 26-35.

Likert, R. A. (1932). A technique for the measurment of attitudes. *Archives of Psychology, 22*, 140-155.

Mattell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert scale items? Reliability and validity. *Educational Measurement, 31*, 657-674.

Menon, G., Raghubir, P., & Schwarz, N. (1995). Behavioral frequency judgements: An accessibility-diagnosticity framework. *Journal of Consumer Research, 22*, 212-228.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practices, 14*, 5-8.

Motl, R. W., McAuley, E., & DiStefano, C. (2005). Is social desirability associated with self-reported physical activity? . *Preventive Medicine, 40*, 735-739.

Noar, S. M., & Zimmerman, R. S. (2005). Health behavior theory and cumulative knowledge regarding health behaviors: are we moving in the right direction? *Health Education Research, 20*, 275-290.

Ogden, J. (2003). Some problems with social cognition models: A pragmatic and conceptual analysis. *Health Psychology, 22*, 424-428.

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 5*, doi:10.1186/1479-5868-1185-1156.

Prochaska, J. O., & DiClemente, C. C. (1982). Transtheoretical therapy: Toward a more integrative model of change. *Psychotherapy: Theory, Research & Practice, 19*, 276-288.

Remmers, H. H., & Ewart, E. (1941). Reliability of multiple,choice measuring intruments as a function of the Spearman-Brown prophecy formula, III. *Journal of Educational Psychology, 32*, 61-66.

Rhodes, R. E., Blanchard, C. M., & Blacklock, R. E. (2008). Do physical activity beliefs differ by age and gender? *Journal of Sport and Exercise Psychology, 30*, 412-423.

Rhodes, R. E., Blanchard, C. M., & Matheson, D. H. (2006). A multi-component model of the theory of planned behavior. *British Journal of Health Psychology, 11*, 119-137.

Rhodes, R. E., Blanchard, C. M., Matheson, D. H., & Coble, J. (2006). Disentangling motivation, intention, and planning in the physical activity domain. *Psychology of Sport and Exercise, 7*, 15-27.

Rhodes, R. E., & Courneya, K. S. (2003a). Investigating multiple components of attitude, subjective norm, and perceived behavioral control: An examination of the theory of planned behavior in the exercise domain. *British Journal of Social Psychology, 42*, 129-146.

Rhodes, R. E., & Courneya, K. S. (2003b). Self-efficacy, controllability, and intention in the theory of planned behavior: Measurement redundancy or causal independence? *Psychology and Health, 18*, 79-91.

Rhodes, R. E., & Courneya, K. S. (2004). Differentiating motivation and control in the theory of planned behavior. *Psychology, Health and Medicine, 9*, 205-215.

Rhodes, R. E., Matheson, D. H., & Blanchard, C. M. (2006). Beyond scale correspondence: A comparison of continuous open scaling and fixed graded scaling when using social cognitive constructs in the exercise domain. *Measurement in Physical Education and Exercise Science, 10*, 13-39.

Rhodes, R. E., & Pfaeffli, L. A. (in press). Mediators of physical activity behavior change among adult non-clinical populations: A review update. *Annals of Behavioral Medicine*.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.

Schwarz, N., Hippler, H. J., Deutsch, B., & Strack , F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly, 49*, 388-395.

Symons Downs, D., & Hausenblas, H. A. (2005). Exercise behavior and the theories of reasoned action and planned behavior: A meta-analytic update. *Journal of Physical Activity and Health, 2*, 76-97.

U.S. Department of Health and Human Services. (2003). Prevalence of physical activity, including lifestyle activities among adults - United States, 2000-2001. *Morbidity & Mortality Weekly Report, 15*, 764-769.

Warburton, D. E. R., Katzmarzyk, P., Rhodes, R. E., & Shephard, R. J. (2007). Evidence-informed physical activity guidelines for Canadian adults. *Applied Physiology, Nutrition and Metabolism, 32*, S16-S68.

Table 1
Demographics of the Sample by Scaling Group

	5L	5S	7L	7S	$F_{3,408}$	*p*
Age yrs *M* (*SD*)	22.17 (6.72)	22.24 (7.52)	22.18 (6.94)	22.09 (5.48)	0.08	1.00
Year in university *M* (*SD*)	1.85 (1.10)	1.90 (1.15)	1.77 (1.19)	1.95 (1.35)	0.43	0.73
Past physical activity frequency	5.00 (3.63)	5.47 (3.01)	5.81 (3.24)	4.97 (3.18)	1.55	0.20
2-week physical activity frequency	4.35 (3.79)	5.25 (3.73)	5.22 (3.43)	4.92 (3.51)	1.43	0.24
					χ^2	*p*-value
Gender (% female)	.63	.68	.71	.68	1.45	0.69
Meeting physical activity guidelines (%)	.68	.73	.71	.68	1.43	0.70

Note: 5L = 5-point Likert, 5S = 5-point Semantic, 7L = 7-point Likert-type, 7S = 7-point Semantic.

Table 2
Means and Variances across Scaling Response Options

		5L	5S	7L	7S	$F_{3,408}$	*p*	Post-Hoc
Affective attitude	*M*	3.85	3.75	5.28	5.27			
	S^2	0.78	0.49	1.98	1.24	14.82	.00	5S < All; 5L< 7S, 7L
Instrumental attitude	*M*	4.64	4.45	6.43	6.32			
	S^2	0.24	0.32	0.91	0.47	7.03	.00	5L < All; 5S < 7L
Subjective norm	*M*	3.75	3.89	5.34	5.56			
	S^2	0.52	0.33	1.21	0.69	12.55	.00	5S, 5L < 7S, 7L
Perceived behavioral control	*M*	4.32	4.04	5.84	5.67			
	S^2	0.59	0.62	1.51	1.15	9.00	.00	5L, 5S < 7S, 7L
Intention	*M*	3.83	3.85	5.36	5.31			
	S^2	0.92	0.63	2.46	1.89	17.29	.00	5S, 5L < 7S, 7L

Note: F = Levene's statistic for homogeneity of variance; post hoc represent Levene's statistic for homogeneity of variance at $p < .05$; 5L = 5-point Likert, 5S = 5-point Semantic, 7L = 7-point Likert-type, 7S = 7-point Semantic. Post-hocs are reported at $p < .05$.

Table 3
Cronbach's Alpha Reliabilities across Response Options

	5L	5S	7L	7S	χ^2	*p*	Post-Hoc
Affective attitude	.88	.70	.91	.73	31.26**	.00	5S, 7S < 5L, 7L
Instrumental attitude	.67	.81	.88	.78	16.47**	.00	5L <5S, 7L; 7S<7L
Subjective norm	.68	.71	.80	.70	0.43	.23	
Perceived behavioral control	.78	.80	.87	.69	11.42**	.01	7S, 5L < 7L
Intention	.91	.88	.95	.91	10.56**	.01	5S, 5L, 7S < 7L

Note: Test statistic and post hocs use the test of independent associations (Hedges & Olkin, 1985); Post-hocs are reported at $p < .05$. 5L = 5-point Likert, 5S = 5-point Semantic, 7L = 7-point Likert-type, 7S = 7-point Semantic.

Table 4.
Theory of Planned Behavior Construct Correlations with Intention across Scaling Response Options

	5L	5S	7L	7S	χ^2	*p*	Post-Hoc
Affective attitude	.72	.62	.83	.58	15.96**	.00	5S, 7S, 5L < 7L
Instrumental attitude	.55	.41	.47	.29	5.47	.14	
Subjective norm	.44	.40	.58	.48	3.08	.37	
Perceived behavioral control	.49	.40	.51	.36	2.23	.51	
Multiple *R*	.79	.71	.85	.67	11.45**	.01	7S, 5S < 7L

Note: Test statistic and post hocs use the test of independent associations (Hedges & Olkin, 1985); Post-hocs are reported at $p < .05$. 5L = 5-point Likert, 5S = 5-point Semantic, 7L = 7-point Likert-type, 7S = 7-point Semantic. Multiple R includes affective attitude, instrumental attitude, subjective norm, and perceived behavioral control as independent variables.

Table 5.

Theory of Planned Behavior Construct Correlations with Behavior across Scaling Response Options

	5L	5S	7L	7S	χ^2	*p*	Post-Hoc
Affective attitude	.52	.33	.45	.44	2.83	.42	
Instrumental attitude	.27	.29	.19	.20	0.83	.84	
Subjective norm	.41	.39	.19	.28	3.82	.28	
Perceived behavioral control	.43	.30	.24	.19	4.16	.24	
Intention	.60	.60	.52	.52	1.35	.71	
Multiple *R*	.62	.60	.52	.52	1.82	.61	

Note: Test statistic uses the test of independent associations (Hedges & Olkin, 1985). 5L = 5-point Likert, 5S = 5-point Semantic, 7L = 7-point Likert-type, 7S = 7-point Semantic. Multiple R includes intention and perceived behavioral control as independent variables.

Table 6.
Multiple Regression of Physical Activity Intention on the Theory of Planned Behavior across Conditions.

Predictors	*R*	R^2	*F*	*df*	*p*	β
5-point Likert	.79	.62	42.42	4,106	.00	
PBC						.18**
Instrumental attitude						.20**
Affective attitude						.50**
Subjective norm						.13
5-point Semantic	.71	.50	23.37	4,94	.00	
PBC						.21**
Instrumental attitude						.13
Affective attitude						.44***
Subjective norm						.23**
7-point Likert	.85	.72	59.46	4,94	.00	
PBC						.08
Instrumental attitude						-.12
Affective attitude						.75**
Subjective norm						.19**
7-point Semantic	.67	.44	18.59	4,93	.00	
PBC						.20*
Instrumental attitude						-.01
Affective attitude						.42**
Subjective norm						.25**

Note. *$p < .05$; **$p < .01$; β = standardized coefficient. *df* = degrees of freedom. PBC = perceived behavioral control.

Table 7.
Regression of Physical Activity Behavior on Intention and Perceived Behavioral Control Across Conditions .

Predictors	*R*	R^2	*F*	*df*	*p*	β
5-point Likert	.62	.38	33.21	2,107	.00	
Intention						.51**
PBC						.18*
5-point Semantic	.60	.36	27.76	2,98	.00	
Intention						.57**
PBC						.08
7-point Likert	.52	.27	17.38	2,96	.00	
Intention						.53**
PBC						-.03
7-point Semantic	.52	.27	17.82	2,97	.00	
Intention						.52**
PBC						-.01

Note. *p < .05; **p < .01; β=standardized regression coefficient. *df* = degrees of freedom. PBC = perceived behavioral control.