Published in final edited form as:

Am J Occup Ther. 2011 ; 65(2): e20–e28. doi:10.5014/ajot.2011.000778.

Psychometric Study of the Occupational Self Assessment with Adolescents After Infectious Mononucleosis

Renee Taylor, Sun Wook Lee, Jessica Kramer, Yukiko Shirashi, and Gary Kielhofner

Abstract

This study examined the psychometric characteristics of the Occupational Self Assessment (OSA) which measures clients' perceptions of their own competence and the value they assign to occupations. Two hundred and ninety-six adolescents with acute mononucleosis completed the OSA along with the Fatigue Scale, the Checklist of Infectious Symptoms, the Child Health Questionnaire, and the Perceived Stress Scale. Results indicated that OSA items coalesced to capture the intended constructs that the rating scales functioned as intended and that over 90% of adolescents were validly measured. The OSA showed adequate sensitivity and was stable over time. As expected, the OSA measures of competence were moderately associated with infectious symptoms, fatigue severity, health status, and stress while the measure of values was not and both measures were not associated with age, gender, or ethnicity. Finally, as expected, adolescents who had not recovered from mononucleosis after 12 months showed lower measures on the competence scale while still attaching the same value/importance to occupational participation when compared to recovered adolescents.

Keywords

Mononucleosis; Occupational Self Assessment; Reliability; Validity

Introduction

This study examined the reliability, sensitivity, and validity of the Occupational Self Assessment (OSA) (Baron, Kielhofner, Iyenger, Goldhammer, & Wolenski, 2006) in a sample of adolescents during their recovery from infectious mononucleosis (MONO). Severe fatigue is common during the acute phase of MONO (Buchwald, Rea, Katon, Russo, & Ashley, 2000; Hickie et al., 2006; White et al., 2001) and becomes chronic in some individuals (Buchwald et al., 2000). Occupational consequences of fatigue in adolescents include prolonged absence from school, disruption of family activities, and loss of contact with peers (Buchwald et al., 2000).

The OSA is a client-centered evaluation tool designed to capture individuals' perceptions of how illness and disability affect their occupations. Based on the Model of Human Occupation (Kielhofner, 2008), the OSA allows clients to rate their own competence and assign value to 21 areas of performance/participation.

A series of studies using item response theory, via the Rasch measurement model (Rasch, 1960/1980; Smith, Wakely, de Kruif, & Swartz, 2003; Wolfe & Chiu, 1999; Wright & Stone, 1979), have examined internal validity of the assessment (i.e., whether the items cohered to capture the intended constructs and whether the hierarchy of items representing less or more of the intended trait was meaningful) and instrument sensitivity. The first study of the OSA provided evidence of internal validity but identified the rating scale as not adequately sensitive (Kielhofner & Forsyth, 2001). Three iterative studies found that a four-point rating scale resulted in an adequately sensitive instrument (Kielhofner, Forsyth,

Kramer, & Iyenger, 2009). Across the studies, items cohered to capture the two measured constructs and the item hierarchy showed a meaningful progression that was similar across the studies.

Another study of 112 individuals demonstrated that the hierarchies of competence and values items were stable across two administrations of the OSA, indicating that the instrument measures the same constructs at both initial and subsequent evaluation (Kielhofner, Dobria, Forsyth, & Kramer, in press). Moreover, the OSA competence scale detected significant change in 32% of respondents, and the values scale detected significant change in 49% of respondents.

Aim

This study continues the psychometric examination of the OSA by focusing on an adolescent population and by combining an item response theory approach via the Rasch Measurement model (Rasch, 1960/1980; Wright & Stone, 1979) with classical test theory approaches to establishing assessment validity. This combined approach enables a more comprehensive assessment of validity and has been applied to the investigation of other self report assessments (Asgari & Kramer, 2008; Kyngdon, 2004; Mok, 2004).

Method

Validity refers to the capacity of an instrument to measure the intended construct (Bensen & Schell, 1997). Both item response theory and classical test theory are concerned with validity. However, they approach the problem of providing empirical evidence related to validity in different ways as noted in the following sections.

Rasch Model Approach

Item response theory (Rasch analysis) generates evidence that describes the underlying characteristic, or “latent trait,” that is intended to be captured by the scale (Wolfe & Smith, 2007). One type of evidence is whether the items effectively capture the intended trait; each item is analyzed as to whether it “fits” with the rest of the items in targeting the trait. To be considered structurally valid, the vast majority of a scale's items should have adequate fit. In previous studies (Kielhofner & Forsyth, 2001; Kielhofner et al., 2009; Kielhofner et al., in press), item fit to each scale provided evidence of good structural validity. Therefore, we expected 90% of OSA items to fit the constructs of occupational competence and value.

Rasch analysis also generates evidence of validity is by examining the continuum less to more of the latent trait formed by the different items. Rasch analysis orders (i.e., calibrates) items on this continuum, allowing determination of whether it makes sense from a theoretical and/or practical perspective. Since previous research found that the OSA competence and values continua are theoretically interpretable in an adult population, we expected them to be similarly ordered in this adolescent population.

The Rasch model can also assess substantive validity which is the extent to which theory explains how and why persons completing the assessment differ in their responses (Wolfe & Smith, 2007). Just as with items, a person “fit” statistic provides an indication of whether or not each person's response pattern fit the underlying item hierarchy (an indication that the person was validly measured). A second concern is whether the participants collectively used the rating scale in the intended manner, as a four-point continuum measuring less to more competence and value. For instance, OSA respondents rate items in terms of their own competence (“I have a lot of problem doing this,” “I have some difficulty doing this,” “I do this well,” or “I do this extremely well”). These ratings are intended to be ordered from less to more and, if not used as intended, threaten the validity of the scale. Finally, critical

questions can be asked about whether the ordering of items and the rating scale are stable over time. We examined this in the present study with a subsample for which we had baseline and one-year follow-up OSA data. We also examined whether the OSA rating scales were able to detect change in this group.

Classical Test Theory Approach

From the perspective of classical test theory, one examines relationships based on theoretical assumptions about the targeted construct and its relationship to other variables. Convergent validity refers to the expectation that measures intended to capture related traits should be correlated. Divergent validity refers to the expectation that measures of unrelated traits should not be correlated (Campbell & Fiske, 1959; Kielhofner, 2006). In this study, we tested whether the correlation of the two OSA variables (competence and value) with several other variables would be as expected based on their theoretical relationships.

Discriminant validity uses groups who are known to differ on some variable(s) relevant to the trait and who, thus, are expected to show differences in the trait (Campbell & Fiske, 1959; Kielhofner, 2006). A scale's validity is tested through asking whether the groups will differ significantly in the measures obtained with the scale. In this study, we compared adolescents who had not recovered from fatiguing symptoms at 12-month follow-up with matched controls who had recovered. We expected that those with persistent fatigue would show lower measures on self-reported competence while still attaching the same value to occupational participation as the non-fatigued controls.

Hypotheses

Nine hypotheses were tested in this study. Each hypothesis, its association with Rasch or classical test theory, the logic of the hypothesis, and the analysis and criteria used to test the hypothesis are shown in Table 1.

Sample—This IRB-approved study was conducted as part of a larger epidemiological study investigating the incidence, correlates, and consequences of severe fatigue in adolescents with MONO (Katz, Shiraishi, Mears, Binns, & Taylor, in press). Three hundred and one adolescents aged 12–18 who were recently diagnosed with acute MONO participated. They were recruited from schools, hospitals, outpatient clinics, and private physician practices. Of this sample, 296 completed both scales of the OSA and all of the other self-report measures required for this study. The follow-up sample was comprised of 31 participants who did not consider themselves to be fully recovered and 59 matched controls from the group who considered themselves fully recovered from MONO at 12 months following initial enrollment and after the acute phase of MONO. Two matched controls were identified for each participant with persistent fatigue; three of the matched controls did not complete the self-report measures necessary for inclusion in this study.

Measures—In addition to the OSA, the following measures were used for this study. The checklist of infectious symptoms (CFS) (Buckwald et al., 2000) is a self-report measure of the presence and severity of MONO symptoms. It was developed for and used in a large-scale adult study of CFS following MONO (Buckwald et al., 2000). The Fatigue Scale (Chalder et al., 1993) which evaluates fatigue severity has been found to be reliable and valid (Chalder et al., 1993) and has been used extensively in a variety of community-based and clinic-based studies of fatigue (e.g., Pawlikowska et al., 1994). The Child Health Questionnaire (CHQ) (Landgraf, Abetz, & Ware, 1996) is a measure of health-related quality of life with well established psychometric properties. For this study, the Global Health Scale (a general rating of the respondent's health) was used. The Perceived Stress Scale (Cohen, Kamarck, & Mermelstein, 1983) measures self-reported stress in community-

based populations; it has well-established psychometric properties and has been used in numerous studies of disabled and non-disabled populations.

Data Collection Approach—Participants completed the OSA and the other measures at home in the presence of a research assistant. Baseline data were gathered during the acute stage of MONO. Follow-up data was collected 12 months following the initial infection with MONO.

Analysis—Hypotheses derived from item response theory were tested using a Rasch rating scale model (Wright & Masters, 1982) via Winsteps version 3.49 (Linacre, 2004). The extent to which items captured the underlying trait was evaluated using item fit statistics. In keeping with recommended standards (Bray, Fisher, & Duran, 2001; Velozo et al., 1999; Wright & Linacre, 1994), fit statistics of MnSq > 1.4 associated with Zstd > 2.0 were taken as indications of an item misfit. The sensitivity of the OSA was evaluated by converting the person separation index into strata, or $(4[\text{separation index}] + 1)/3$, which indicate the number of significant different levels of the construct represented by the items (Wright & Masters, 1982).

The rating scales were examined to determine if clients were using them as intended as evidenced by incremental increase in the average client measures associated with each rating category and in the threshold difficulties of each rating scale, and by outfit MnSq fit statistics for each category less than 2.0 (Linacre, 2002).

The extent to which clients were validly measured was explored by determining the percentage that had acceptable fit statistics on baseline and 12-month responses to the OSA. Client fit statistics of MnSq > 1.4 associated with Zstd > 2.0 were taken as indications of a misfit.

When analyzing the responses of individuals over time, one must distinguish changes in estimated measures related to actual change in the intended construct from that due to measurement error (Wolfe & Chiu, 1999). In self-report test-retest situations, if participants complete the assessment the same way each time, one can have much higher confidence that changes in measures are due to actual changes in the underlying trait as opposed to measurement error. To evaluate the stability of the OSA rating scales and items over time, we followed the procedure described by Wolfe and Chiu (1999). First, each time point for the 90 participants on whom we had repeated measures was analyzed separately using a Rating Scale Model (RSM) approach. Then resulting item and rating scale Rasch estimates from baseline and 12 months were compared for significant differences over time as given by:

$$\frac{\text{Time 1 item measure} - \text{Time 2 item measure}}{\sqrt{(\text{Time 1 item measure error})^2 + (\text{Time 2 item measure error})^2}}$$

$$\sqrt{(\text{Time 1 item measure error})^2 + (\text{Time 2 item measure error})^2}$$

Values greater than ± 2 are taken as indications of significant differences in assessment parameters between administrations. In this instance, steps must be taken to scale readministrations onto a common metric and adjust person measures accordingly; this ensures a rigorous evaluation of reported change over time. In this study, because rating scale steps showed small but significant changes, initial and 12-month OSA client measures were obtained using a common rating scale structure derived from a combined, simultaneous

analysis of both time points. Finally in order to examine the ability of the OSA scales to detect change in individuals, we examined the numbers of adolescents who showed significant change from baseline to one year follow-up. For each individual, a change score was derived from following formula which accounts for measurement error associated with each estimated measure:

$$\frac{\text{Time 1 adolescent measure} - \text{Time 2 adolescent measure}}{\sqrt{(\text{Time 1 adolescent measure error})^2 + (\text{Time 2 adolescent measure error})^2}}$$

Values greater than ± 2 indicate significant differences between administrations.

Hypotheses derived from classical psychometric theory were tested using t-tests, Pearson's correlation analyses, and ANOVA tests. Because of multiple comparisons, a conservative alpha level ($p < 0.01$) was used to determine significance.

Results

Hypothesis I: The OSA items will coalesce to capture the intended constructs of occupational competence and values of adolescents with MONO

At baseline, 90.48% of the occupational competence scale items and 90.48% of the values scale items met fit requirements (see Table 2). Overall, the findings supported the hypothesis the scale items would coalesce to capture the intended constructs.

Hypothesis II: The occupational competence and value rating scales will function as intended among adolescents with MONO

At baseline, the occupational competence rating scale was used as intended as average client measures and thresholds were ordered and all rating scales had acceptable fit statistics (see Table 3). The values rating scale was also used as intended at baseline (see Table 3).

Hypothesis III: Ninety percent of all adolescents will be validly measured by the OSA both at baseline and 12 months

Fit statistics could not be calculated on small numbers of participants who used the highest ratings for all items; percentages are based on the total participants for whom fit statistics could be calculated. At baseline, 7.12% misfit on the competence scale and 8.87% misfit on the values scale. At 12 months clients, 8.23%, misfit on the competence scale and 6.74% misfit on the values scale. Thus, in all instances over 90% met criteria for being validly measured.

Hypothesis IV: The OSA rating scales will show adequate sensitivity to discriminate between adolescents with different levels of occupational competence and value for their occupation at baseline

At baseline, the OSA competence items measured almost four significantly different levels (3.99 strata) of competence among adolescents (person separation index = 2.74). The OSA value items also measured four significantly different levels (4.03 strata) of value among adolescents (person separation index = 2.77).

Hypothesis V: OSA occupational competence and value rating scales and the hierarchy of each scale will not be substantially different over time

The OSA had evidence of minimal variance between baseline and 12 months. First, the competence item calibrations remained stable over time (Table 4). Only one competence item became significantly more challenging at 12 months. The OSA competence rating scale thresholds were not all stable over time; adolescents felt more competent and had an easier time using the highest rating category of “extremely well” at baseline (Table 5). However, the estimates were only separated by 0.44 logits and the associated errors were small. That is, the difference was significant, yet small.

The majority of the items on the values item hierarchy were the same over time (Table 6). Three OSA values items were significantly different at 12 months. The OSA value rating scale thresholds were not all stable over time; adolescents indicated more value for more items and had an easier time using the highest rating category of “most important” at baseline (Table 7). However, as with the competence rating scale, the baseline and 12-month parameter estimates differed by a small amount (0.29 logits) and the associated errors were small. That is, the difference was significant, yet small.

To account for changes in the rating scale over time, initial and 12 month OSA client measures were obtained using a common rating scale structure derived from a combined, simultaneous analysis of both time points. Although four items varied over time, no item varied over time on both scales. Therefore, the OSA items were considered invariant over time and considered to also have a shared item estimate at both time points. Deriving person measures from this common rating scale places the baseline and 12-month assessment on a shared and common metric, or frame of reference, and derives a more conservative but more rigorous estimates of change (Wolfe & Chiu, 1999). These adjusted baseline and 12 month measures were used to answer hypotheses 7 and 8.

Hypothesis VI: The OSA competence and values scales will detect significant change in 30% of the adolescents

The OSA competence scale detected significant change in 34% of the participants while the values scale detected significant change in 38%. The OSA competence score detected significant increases in 34% of adolescents who were recovered and in 19% of those with ongoing fatigue. The OSA competence score detected significant decreases in 3% of those recovered and 10% of those with ongoing fatigue. The OSA values scale detected similar rates of significant increases and decreases in both groups. Twenty-four percent of recovered and 26% of those with ongoing fatigue showed increases, while 14% of those recovered and 13% of those with ongoing fatigue showed decreases.

Hypothesis VII: Occupational competence and value measures among adolescents with IM will not be associated with their age, gender, or ethnicity

There were no significant associations between the adjusted OSA competence and OSA value measure with gender and ethnicity at baseline as shown in Tables 8 and 9. There were weak or no associations between OSA value baseline measure and age, as shown in Table 10.

Hypothesis VIII: Client occupational competence measures will show significant associations with measures of symptoms, fatigue, health status, and stress while value measures will not

Table 11 shows the correlations between the adjusted OSA measures and the measures of infectious symptoms, fatigue, health status, and stress. As anticipated, the OSA competence scale was moderately correlated with infectious symptoms, fatigue, health status, and

perceived stress. As expected, the OSA value measures were not correlated with infectious symptoms or fatigue severity. Significant but minimal correlations were found with stress and global health.

Hypothesis IX: Non-recovered adolescents will show lower measures on self-reported competence while still attaching the same value to occupational participation as recovered adolescents

T-tests comparing adolescents who were non-recovered at 12 months follow-up with matched controls who were recovered adolescents revealed that non-recovered adolescents had lower adjusted competence measures ($M = 54.98$, $SD = 8.79$) than those who were recovered ($M = 62.21$, $SD = 8.33$), $t(88) = -3.84$, $p = 0.00$. Moreover, as hypothesized, there was no significant difference between these groups in the reported values/value measures.

Discussion

This psychometric study provided evidence of the validity, reliability, sensitivity, and stability of OSA in a sample of adolescents. First, OSA items coalesced to capture the intended constructs. Findings support use of the OSA as a valid tool to measure what adolescents perceive as areas of competence and what they perceive as important in terms of their everyday activities. All of the competence items meaningfully cohered to measure the construct of competence in the sample, and all of the value items that are in the OSA cohered to measure the construct of values.

Additionally, findings indicated that the rating scales of the OSA were used as intended, as continuums representing less to more competence and values. The adolescents appropriately used the higher rating categories to indicate more competence and value, and the lower rating categories to indicate less competence and less value.

Moreover, over 90% of the adolescents were validly measured and gave responses consistent with Rasch model expectations. A typical criterion applied to person misfit is that no more than 5% of the respondents should have poor fit (Smith, Conrad, Chang, & Piazza, 2002). However, 10% has been considered an acceptable target for the OSA, as the self report is completed without the type of extensive training associated with professionally administered assessments (Kielhofner et al., 2009). Identifying clients whose pattern of responses misfit can be done using the keyforms found in the OSA manual (Baron, et al., 2006) and should be followed with a clarifying discussion to make sure the clients responses are understood.

Another indicator of psychometric stability was that the competence and value items had the sensitivity to measure approximately four significantly different levels of competence and value among adolescents. This finding supports the use of the OSA as a tool for identifying group differences in reported competence and value.

Lastly, this study illustrated the use of a Rasch-based procedure (Wolfe & Chiu, 1999) to examine the stability of OSA items and rating scales over time when used by adolescents with MONO. Results suggest that the relative difficulty and value of everyday tasks does not change over time for this group, as given by the relative stability of item measures over time. However, over time, this group is slightly more selective in the extent to which they report the highest level of competence and value for everyday activities as given by the significant differences in the highest rating scale thresholds over time. While statistically significant, the difference in rating scale use over time was small and may be considered negligible for use in practice.

The OSA competence scale detected significant change in approximately a third of the adolescents from baseline to one year later. As expected, recovered adolescents showed a higher rate of increase and a lower rate of decrease in competence scores than adolescents with ongoing fatigue. The OSA values scale detected significant change in about a quarter of both groups showed increases, while 13% with ongoing fatigue and 14% of those recovered had significantly decreased scores. The values change scores likely reflect a combination of developmental changes and the impact of an illness experience on adolescents' thoughts on what was important. That the values scale detected somewhat more change than the competence scale is consistent with previous research on the OSA (Kielhofner et al., in press). Overall, the findings indicate that the OSA can detect statistically significant and meaningful change in individuals.

The OSA competence scale was associated with measures of symptoms, fatigue, health status, and stress as expected. Moreover, as expected, the OSA values scale was not associated with these measures. Finally, there was no association between adolescents' perceived competence and values and the sociodemographic variables of age, gender, or ethnicity. These findings provide evidence of convergent and divergent validity in the OSA.

Twelve months after infection, when non-recovered adolescents were compared with adolescents who recovered, the former adolescents showed lower measures on the competence scale but not the values scale. This finding was expected because we hypothesized that while competence would be affected by prolonged fatigue, values would be likely to be stable over time. This finding provides evidence of discriminant validity.

Conclusion

Overall, findings from this study contributed additional evidence of the OSA's reliability and validity from both an item response and classical test theory approach. Findings also indicate that the OSA is an adequately sensitive measure capable of discriminating between clients with different competence and values and of detecting meaningful change. Finally, the study provided evidence that the OSA can be used in a valid and reliable manner with adolescents.

Acknowledgments

We thank the Chicago Children's Memorial Hospital's Pediatric Practice Research Group and all of the physicians and school nurses who referred patients into the study. We also thank student volunteers who contributed to the recruitment of physicians and school nurses.

References

- Asgari A, Kramer J. Construct validity and factor structure of the Persian Occupational Self Assessment (OSA) with Iranian students. *Occupational Therapy in Health Care*. 2008; 22(2/3):187–200.
- Baron, K.; Kielhofner, G.; Iyenger, A.; Goldhammer, V.; Wolenski, J. *The Occupational Self Assessment (version 2.2)*. Model of Human Occupation Clearinghouse, Department of Occupational Therapy, College of Applied Health Sciences, University of Illinois at Chicago; Chicago: 2006.
- Benson, J.; Schell, BA. Measurement theory: Application to occupational and physical therapy. In: Van Deusen, J.; Brunt, D., editors. *Assessment in occupational therapy and physical therapy*. W.B. Saunders; Philadelphia, PA: 1997. p. 3-24.
- Bray K, Fisher AG, Duran L. The validity of adding new tasks to the assessment of motor and process skills. *American Journal of Occupational Therapy*. 2001; 55(4):409–415. [PubMed: 11723985]

- Buchwald DS, Rea TD, Katon WJ, Russo JE, Ashley RL. Acute infectious mononucleosis: Characteristics of patients who report failure to recover. *American Journal of Medicine*. 2000; 109(7):531–537. [PubMed: 11063953]
- Campbell DT, Fiske DW. Convergent and discriminant validation for the multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56:81–105. [PubMed: 13634291]
- Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, et al. Development of a fatigue scale. *Journal of Psychosomatic Medicine*. 1993; 37:147–153.
- Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of Health and Social Behavior*. 1983; 24:385–396. [PubMed: 6668417]
- Hickie I, Davenport T, Wakefield D, Vollmer-Conna U, Cameron B, Vernon SD, et al. Post-infective and chronic fatigue syndromes precipitated by viral and non-viral pathogens: Prospective cohort study. *BMJ*. 2006; 333(7568):575. [PubMed: 16950834]
- Katz BZ, Shiraishi Y, Mears C, Binns H, Taylor RR. Chronic fatigue syndrome following infectious mononucleosis in adolescents: A prospective cohort study. *Pediatrics*. in press.
- Kielhofner, G. Developing and evaluating quantitative data collection instruments. In: Kielhofner, G., editor. *Research in occupational therapy: Methods of inquiry for enhancing practice*. F.A. Davis; Philadelphia: 2006. p. 169
- Kielhofner, G. *Model of human occupation: Theory and application*. 4th ed.. Lippincott, Williams, & Wilkins; Baltimore: 2008.
- Kielhofner G, Dobria L, Forsyth K, Kramer J. The Occupational Self Assessment (OSA): Stability and the ability to detect change over time. *OTJR: Occupation, Participation and Health*. in press.
- Kielhofner G, Forsyth K. Measurement properties of a client self-report for treatment planning and documenting therapy outcomes. *Scandinavian Journal of Occupational Therapy*. 2001; 8(3):131–139.
- Kielhofner G, Forsyth K, Kramer J, Iyenger A. Developing the Occupational Self Assessment: The use of Rasch analysis to assure internal validity, sensitivity, and reliability. *British Journal of Occupational Therapy*. 2009; 72(3):94–104.
- Kyngdon A. Comparing factor analysis and the Rasch model for ordered response categories: An investigation of the scale of gambling choices. *Journal of Applied Measurement*. 2004; 5:398–418. [PubMed: 15496747]
- Landgraf, JL.; Abetz, L.; Ware, JE. *The CHQ User's Manual*. The Health Institute, New England Medical Center; Boston: 1996.
- Linacre JM. Optimizing rating scale category effectiveness. *Journal of Applied Measurement*. 2002; 3:85–106. [PubMed: 11997586]
- Linacre, JM. *Winsteps (Version 3.49)*. 2004. Author
- Mok MMC. Validation of scores from self-learning scales for primary students using true-score and Rasch measurement methods. *Journal of Applied Measurement*. 2004; 5:258–286. [PubMed: 15243173]
- Pawlikowska T, Chalder T, Wessely S, Wright D, Hirsch S, Wallace P. A population based study of fatigue and psychological distress. *British Medical Journal*. 1994; 308:763–766. [PubMed: 7908238]
- Rasch, G. *Probabilistic models for some intelligence and attainment tests*. expanded edition. Danish Institute for Educational Research; The University of Chicago Press; Copenhagen: Chicago: 1960/1980. (1980) with foreword and afterword by B.D. Wright
- Smith EV Jr, Conrad KM, Chang K, Piazza J. An introduction to Rasch measurement for scale development and person assessment. *Journal of Nursing Measurement*. 2002; 10(3):189–206. [PubMed: 12885145]
- Smith EV Jr, Wakely M, de Kruif R, Swartz C. Optimizing ratings scales for self-efficacy (and other) research. *Educational and Psychological Measurement*. 2003; 63:369–391.
- Veloza C, Kielhofner G, Gern A, Lin F-L, Azhar F, Lai JS, et al. Worker Role Interview: Toward validation of a psychosocial work-related measure. *Journal of Occupational Rehabilitation*. 1999; 9(3):153–168.

- White PD, Thomas JM, Kangro HO, Bruce-Jones WDA, Amess J, Crawford DH, et al. Predictions and associations of fatigue syndromes and mood disorders that occur after infectious mononucleosis. *Lancet*. 2001; 358(9297):1946–1954. [PubMed: 11747919]
- Wolfe EW, Chiu CWT. Measuring change across multiple occasions using the Rasch Rating Scale Model. *Journal of Outcome Measurement*. 1999; 3(4):360–381. [PubMed: 10572387]
- Wolfe EW, Smith EV Jr. Understanding Rasch measurement: Instrument development tools and activities for measure validation using Rasch models: Part II-Validation activities. *Journal of Applied Measurement*. 2007; 8(2):204–234. [PubMed: 17440262]
- Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Measurement Transactions*. 1994; 8(3):370.
- Wright, BD.; Masters, GN. Rating scale analysis. MESA Press; Chicago: 1982.
- Wright, BD.; Stone, MH. Best Test Design. MESA Press; Chicago: 1979.

Table 1

Study Hypotheses

Hypothesis	Logic	Approach	Evidence-Criteria
1. OSA items will coalesce to capture constructs of competence and values.	Findings should replicate previous findings that OSA items cohered in capturing the latent traits of competence and values (Kielhofner & Forsyth, 2001; Kielhofner et al., 2009; Kielhofner et al., in press).	Rasch analysis	Ninety percent of items have acceptable "fit" to the Rasch model (i.e., MNSQ statistic <1.4 associated with a ZSTD<2).
2. The OSA rating scales will function as intended.	Rating scales should be used by respondents as they were designed to be used; this means that the four-point scale as used by the OSA should demonstrate the intended ordering from less to more competence and value.	Rasch analysis	Average client measures associated with each rating category and threshold difficulties increase incrementally, and outfit MnSq fit statistics for each category are less than 2.0 (Linacre, 2002).
3. Over 90% of all clients are validly measured by the OSA at baseline and 12 months.	For a scale to be considered valid, the vast majority of items should have adequate fit.	Rasch analysis	Ninety of persons will have acceptable "fit" to the Rasch model (i.e., MNSQ statistic <1.4 associated with a ZSTD<2).
4. OSA has the sensitivity to detect at least two significantly different levels of competence and value among respondents at baseline.	The revised OSA items measured six significantly different levels of competence and value in previous research (Kielhofner et al., 2009).	Rasch analysis	At least two strata will be present when the person separation index is converted into strata, using the formula $(4[\text{separation index}] + 1)/3$
5. OSA occupational competence and value rating scales and OSA items on each scale will not be significantly different over time.	Item hierarchy should not differ significantly over time and should remain stable (Kielhofner et al., 2009; Kielhofner et al., in press).	Rasch analysis	Item measures and rating scale threshold difficulties from separate initial and baseline Rasch analysis will not be significantly different.
6. The OSA competence and values scales will detect change in at least a quarter of the adolescents.	In a previous study, 32% and 49% of clients had significantly different measures at time two on competence and values scales, respectively (Kielhofner et al., in press)		OSA values and competence scales will detect statistically significant increases/decreases in 30% of adolescents. Patterns of detected change will make sense for the recovered and chronically fatigued groups.
7. Client competence and value measures will not be associated with age, gender, or ethnicity.	Tests designed to measure different traits are expected to show patterns of association that discriminate between the traits.	Independent samples t-test, one-way ANOVA, and Pearson's correlation tests	No significant difference in OSA measures between groups for age, gender, and ethnicity.
8. Client competence measures will be moderately associated with measures of symptoms, fatigue, health status, and stress, while value measures will not.	Since fatigue affects performance it should be correlated with clients' perception of their own competence but not correlated with the value they attach to areas of performance.	Pearson's correlation	Moderate relationships between competence and symptoms, fatigue, health status, and stress. Negligible/non-significant relationships between OSA values and these variables.
9. Adolescents who have not recovered will show lower measures on self-reported competence while still attaching the same value to occupational participation as recovered adolescents.	Adolescents with persistent fatigue should show lower self-report of competence, but attach similar value to areas of performance when compared to adolescents without fatigue.	T-test	Difference in competence measure between adolescents in non-recovered group and recovered adolescents at 12 months.

Table 2

OSA Competence and Values Item Fit Statistics at Baseline (N = 296)

OSA item	Competence						Values					
	Infit MnSq	Infit Zstd	Outfit MnSq	Outfit Zstd	Infit MnSq	Outfit Zstd	Infit Zstd	Outfit MnSq	Outfit Zstd	Infit Zstd	Outfit MnSq	Outfit Zstd
Concentrating on my tasks	.92	1.00	.93	.80	.82	-.82	-2.50	.82	-2.30			
Physically doing what I need to do	1.04	.40	1.04	.40	.97	-.40	-.40	.96	-.50			
Taking care of the place where I live	1.18	2.00	1.18	1.90	1.42*	4.30*	4.30*	1.40	4.10			
Taking care of myself	.96	-.60	.94	-.60	.93	-1.00	-.90	.90	-1.00			
Taking care of others for whom I am responsible	.77	-3.00	.76	-2.90	1.10	1.30	1.07	.70				
Getting where I need to go	1.01	.20	1.00	.00	.82	-2.40	.82	-.240				
Managing my finances	1.31	2.80	1.35	3.20	1.45*	4.00*	4.00*	1.43*	3.80*			
Managing my basic needs (food, medicine)	.91	-1.10 5.80*	.86	-1.60	1.02	.30	.97	-.30				
Expressing myself to others	1.58*		1.55*	5.50*	1.29	3.40	1.28	3.1				
Getting along with others	1.28	3.20	1.21	2.10	1.04	.60	1.01	.10				
Identifying and solving problems	.90	-1.20 5.20*	.91	-1.10	.90	-1.40	.89	-1.40				
Relaxing and enjoying myself	1.50*		1.56*	5.60*	1.27	3.40	1.25	2.90				
Getting done what I need to do	.78	-2.80	.77	-2.90	.79	-3.00	.78	-2.90				
Having a satisfying routine	.82	-2.30	.81	-2.40	1.21	2.40	1.16	1.80				
Handling my responsibilities	.69	-4.20	.70	-4.00	.84	-2.30	.80	-2.60				
Being involved as a student, worker, volunteer, and/or family member	1.05	.60	1.06	.70	1.12	1.60	1.16	1.90				
Doing activities I like	1.27	3.10	1.22	2.30	1.06	.80	1.12	1.30				
Working towards my goals	.72	-3.70	.71	-3.80	.81	-2.70	.78	-2.70				
Making decisions based on what I think is important	.99	-.10	.95	-.60	.79	-3.10	.77	-3.10				
Accomplishing what I set out to do	.72	-3.60	.73	-3.50	.76	-3.60	.72	-3.80				
Effectively using my abilities	.72	-3.70	.72	-3.70	.78	-3.20	.74	-3.50				

* Item exceeds Rasch fit criteria with MnSq > 1.4 associated with Zstd > 2.

Table 3

Occupational Competence and Values Rating Scale Function at Baseline (N = 296)

Rating scale	Average client measure	Threshold difficulties	Outfit MnSq
Competence			
I have a lot of problem doing this	-.97	(NONE)	1.09
I have some difficulty doing this	.17	-2.35	1.00
I do this well	1.34	-.37	.92
I do this extremely well	2.61	2.72	1.00
Values			
This is not so important to me	-.74	(NONE)	1.09
This is important to me	-.01	-2.20	.95
This is more important to me	.86	.56	.91
This is most important to me	1.80	1.63	1.01

Table 4
OSA Competence Item Measures at Baseline and at 12 Months and Standardized Difference

OSA competence item	Baseline (N = 296)		12 month (N = 90)		Standardized difference
	Measure**	Error	Measure**	Error	
Concentrating on my tasks	1.17	.10	1.23	.20	-.27
Physically doing what I need to do	.07	.10	-.12	.21	.82
Taking care of the place where I live	.75	.10	.79	.20	-.18
Taking care of myself	-1.12	.11	-1.04	.22	-.33
Taking care of others for whom I am responsible	-.73	.11	-1.12	.22	1.59
Getting where I need to go	-.42	.11	-.65	.21	.97
Managing my finances	.79	.12	1.11	.22	-1.28
Managing my basic needs (food, medicine)	-.74	.11	-.47	.21	-1.14
Expressing myself to others	.21	.10	.34	.21	-.56
Getting along with others	-1.15	.11	-1.35	.22	.81
Identifying and solving problems	.01	.10	.13	.21	-.52
Relaxing and enjoying myself	-.25	.10	.37	.21	-2.67*
Getting done what I need to do	.77	.10	.79	.20	-.09
Having a satisfying routine	.81	.10	1.11	.20	-1.34
Handling my responsibilities	.53	.10	.47	.20	.27
Being involved as a student, worker, volunteer, and/or family member	-.22	.10	-.30	.21	.34
Doing activities I like	-.76	.11	-.88	.21	.51
Working towards my goals	.10	.10	-.34	.21	1.89
Making decisions based on what I think is important	-.18	.10	-.07	.21	-.47
Accomplishing what I set out to do	.30	.10	.26	.21	.17
Effectively using my abilities	.04	.10	-.28	.21	1.38

* There is a significant difference between item measures at baseline and at 12 months.

** Item with higher measures are more difficult to perform competently, or more challenging. Items with lower measures are easier to perform competently.

Table 5

OSA Competence Rating Scale Thresholds at Baseline and at 12 Months and Standardized Difference

OSA competence rating scale category	Baseline (N = 296)		12 month (N = 90)		Standardized difference
	Threshold	Error	Threshold	Error	
I have a lot of problem doing this	(NONE) -2.35	(NONE) .09	(NONE) -2.70	(NONE) .20	(NONE) 1.60
I have some difficulty doing this	-0.37	.04	-0.46	.08	1.01
I do this well	2.72	.03	3.16	.06	-4.62*

* Significant difference between baseline and 12 months as indicated by value greater than ± 2.0 .

Table 6

OSA Value Item Measures at Baseline and at 12 Months, and Standardized Difference

OSA value item	Baseline (N = 296)		12 month (N = 90)		Standardized difference
	Measure**	Error	Measure**	Error	
Concentrating on my tasks	.47	.09	.59	.16	-.65
Physically doing what I need to do	.10	.08	.13	.16	-.17
Taking care of the place where I live	1.44	.09	.88	.16	3.05*
Taking care of myself	-1.18	.09	-1.20	.17	.10
Taking care of others for whom I am responsible	-.71	.09	-1.03	.17	1.66
Getting where I need to go	.47	.09	.36	.16	.60
Managing my finances	1.12	.11	.50	.17	3.06*
Managing my basic needs (food, medicine)	-.09	.09	.00	.16	-.49
Expressing myself to others	.27	.08	.46	.16	-1.06
Getting along with others	-.52	.09	-.28	.16	-1.31
Identifying and solving problems	.36	.09	.67	.16	-1.69
Relaxing and enjoying myself	-.32	.08	.23	.16	-3.07*
Getting done what I need to do	-.21	.08	-.30	.16	.50
Having a satisfying routine	.99	.09	1.12	.17	-.68
Handling my responsibilities	.01	.08	-.25	.16	1.45
Being involved as a student, worker, volunteer, and/or family member	-.31	.08	-.05	.16	-1.45
Doing activities I like	-.66	.09	-.48	.16	-.98
Working towards my goals	-.65	.09	-.72	.16	.38
Making decisions based on what I think is important	-.31	.08	-.20	.16	-.61
Accomplishing what I set out to do	-.27	.08	-.41	.16	.78
Effectively using my abilities	.00	.08	.00	.16	.00

* There is a significant difference between item measures at baseline and at 12 months.

** Items with higher measures are more difficult to value, or less important. Items with lower measures are easier to value, or more important.

Table 7

OSA Values Rating Scale Thresholds at Baseline and at 12 Months and Standardized Difference

OSA values rating scale category	Baseline (N = 296)		12 month (N = 90)		Standardized difference
	Threshold	Error	Threshold	Error	
This is not so important to me	(NONE)	(NONE)	(NONE)	(NONE)	(NONE)
This is important to me	-2.20	.06	-2.40	0.13	1.40
This is more important to me	0.56	.03	0.49	.06	1.04
This is most important to me	1.63	.04	1.92	.06	-4.02*

* Significant difference between at baseline and at 12 month as indicated by value greater than ± 2.0 .

Table 8

T-Test Results among OSA Competence and Value by Gender (N = 296)

Variable by gender	N	Mean	SD	t	df	Sig.
Competence						
Male	96	57.26	5.44	.25	245.89	.81
Female	200	57.08	7.39			
Value						
Male	96	52.69	5.91	-1.89	294	.06
Female	200	54.15	6.37			

Table 9

One-way ANOVA Results among OSA Competence and Value upon Ethnicity at Baseline (N = 296)

Variable by ethnicity	Mean	SD	F	Sig.
Competence				
African American	54.96	4.84	.73	.60
Asian	51.76	0		
Multiracial	55.58	4.72		
Latino	55.88	7.02		
White	57.31	6.90		
Other	49.16	0		
Value				
African American	52.83	4.88	.26	.94
Asian	57.31	0		
Multiracial	53.51	2.71		
Latino	52.01	6.22		
White	53.70	6.31		
Other	55.64	7.10		

Table 10

Pearson's Correlation Results on Age and OSA Competence and Value Measure at Baseline (N = 296)

Variable	OSA competence		OSA value	
	R	Significance level	R	Significance level
Age	.05	.38	.17*	.00

*
p<0.01(2-tailed)

Table 11

Pearson Correlation Coefficients between OSA Scales and Infectious Symptoms, Fatigue Severity, Global Health, and Stress (N =296)

Variable	OSA competence		OSA value	
	R	Significance Level	R	Significance Level
Infectious symptoms	-.230*	.00	-.035	.55
Fatigue severity	-.323*	.00	.016	.78
Global health	.287*	.00	.125	.03
Stress for last month	-.558*	.00	-.123	.04

*
p<0.01