BRIEF REPORT

Reliability and Sensitivity to Change of Goal Attainment Scaling in Occupational Therapy Nonclassroom Educational Experiences

Jeanette Koski, Lorie Gage Richards

MeSH TERMS
- clinical clerkship
- clinical competence
- education, graduate
- educational measurement
- occupational therapy

Occupational therapy programs are charged with measuring student progress in nonclassroom experiential components (American Occupational Therapy Association, 2012). Currently, the major nonclassroom educational experience is Level II fieldwork. Level II performance is assessed using the Fieldwork Performance Evaluation, which is inappropriate for measuring doctoral-level experiential component achievement. This study’s purpose was to determine test–retest reliability and sensitivity of Goal Attainment Scaling (GAS) for assessing skill development in nonclassroom occupational therapy experiences. GAS demonstrated high test–retest reliability for each of the five goals, and we found a significant amount of change on the GAS, with higher scores at the 12-wk than at the 6-wk assessment. Results indicate that the GAS is reliable and sensitive to changes in student performance on Level II fieldwork and may therefore warrant investigation as a valid tool to measure student performance in the entry-level doctoral experiential component.


Jeanette Koski, OTD, OTR/L, is Assistant Professor and Academic Fieldwork Coordinator, Division of Occupational Therapy, College of Health, University of Utah, Salt Lake City; Jeanette.koski@hsc.utah.edu

Lorie Gage Richards, PhD, OTR/L, FAHA, is Chair and Associate Professor, Division of Occupational Therapy, College of Health, University of Utah, Salt Lake City.

Occupational therapy in the United States is responding to the demands of a changing health care paradigm by upgrading educational programs from entry-level master’s to entry-level doctoral (OTD) degrees. The intent is to train occupational therapists as health care leaders and consumer advocates (American Occupational Therapy Association [AOTA], 2012). One requirement for OTD programs delineated by the Accreditation Council for Occupational Therapy Education (ACOTE®) is to determine whether a student meets the goals of an in-depth experiential component (AOTA, 2012). The OTD experiential component is designed to “develop occupational therapists with advanced skills” (AOTA, 2012, p. 64) and is to be measured by a finished project and the student’s achievement of “individualized specific objectives” (AOTA, 2012, p. S64) in one of the following eight areas: clinical practice skills, research skills, administration, leadership, program and policy development, advocacy, education, and theory development (AOTA, 2012). An OTD student will set individualized goals in one or more of the focus areas and designate the site at which he or she will develop these skills.

AOTA (2012) has further stipulated that academic programs must have means to objectively assess student performance during and at the completion of the experiential component. Nonclassroom educational experiences in allied health professions represent learning in the natural practice environment and include the development of co-constructed new knowledge, clinical skill development, and application of professional judgment. Assessing competence can include direct observation, chart reviews, and written tests (Salvatori, 1996). Assessments performed in the context of authentic practice, that is, direct observation at the practice site, have a “positive impact on learning as FW [fieldwork] educators value their relevance to students’ future practice” (Smith & Cardillo, 1994, p. 257). The only established assessment for nonclassroom occupational therapy educational experiences is the Fieldwork Performance Evaluation.
The FWPE is used to measure student progress and competence in Level II fieldwork, but the items on the FWPE were not designed to be used for the OTD experiential component, which has a different objective. The goal of Level II fieldwork is to measure entry-level competence (AOTA, 2012; Adler, 2003), which contrasts with the advanced skills goal of the OTD experiential component (AOTA, 2012).

Goal Attainment Scaling (GAS) may be a useful method for assessing competence in the OTD experiential component because it allows for quantitative measurement of individualized goals and can be used in a variety of settings (Doig, Fleming, Kuipers, & Cornwell, 2010; Heavin, Lee-Merrow, & Lewis, 1982; Ottenbacher & Cusick, 1990; Ruble, McGrew, & Toland, 2012). GAS has been studied in mental health, medical, and kindergarten through 12th-grade educational settings (Barnett et al., 1999; Cardillo & Smith, 1994; Doig et al., 2010; Epstein & Hundert, 2002; Joseph et al., 2014; King, McDougall, Palisano, Gritzan, & Tucker, 1999; Kiresuk & Sherman, 1968; Malec, 1999; McCallister, Lincoln, Ferguson, & McCallister, 2010; McGaghtie & Menges, 1975; Turner-Stokes & Williams, 2010). GAS has been shown to have good face and content validity and sensitivity (Malec, 1999; McGaghtie & Menges, 1975) to measure individualized achievement of goals in these settings and, therefore, it is reasonable to study psychometric properties of this measure for occupational therapy student goal achievement in nonclassroom environments.

Currently, only one common extended nonclassroom learning experience exists in occupational therapy education: Level II fieldwork. Establishing test–retest reliability and sensitivity to a change in GAS methodology in assessing Level II fieldwork outcomes may provide insight into whether GAS might be a useful OTD experiential performance assessment methodology for two reasons. First, although Level II fieldwork is different from the OTD experiential component, both have similar key factors. For example, for most Level II fieldwork students and OTD experiential students, supervision will commonly be by people other than program faculty. Additional similarities include a number of student performance constructs that are likely to be assessed in both experiences: professional reasoning, verbal communication, written communication (documentation), and building rapport with colleagues and clients, participants, consumers, and so forth (AOTA, 2012). Second, determining the sensitivity of a measure requires knowledge that a change in skills actually occurred over the course of the experience. The Level II fieldwork experience is the only experience for which there is an established assessment tool that would allow determination of change. Thus, it seems plausible to use the Level II fieldwork experience as an initial test of the reliability and sensitivity to a change in GAS in nonclassroom occupational therapy learning experiences.

The purpose of this study was to examine the test–retest reliability and sensitivity to change of GAS for a nonclassroom educational experience for occupational therapy students to lay the groundwork for its use in evaluating student goal achievement in the experiential component of entry-level OTD programs. The research questions for this study were as follows: (1) Does the score on GAS indicate statistically significant test–retest reliability and (2) is GAS sensitive enough to measure change in occupational therapy student performance over time?

Method

Participants

Participants were occupational therapy students completing the first of two Level II fieldwork assignments in a master’s program and fieldwork educators who rated the students’ goal achievement on both the FWPE and GAS.

Outcome Measures

GAS is a technique used to evaluate the achievement of individualized goals with an ordinal scale. GAS goals are set on the basis of the expected level of function, or the most likely outcome, for each individualized goal (Ottenbacher & Cusick, 1990). In this study, the expected level of function (scored 0) was the individual goal or skill attainment that would be expected for the student to pass fieldwork on that item at each site. Goals were then operationalized so that they represented the most favorable outcome (2), greater than expected outcome (1), less than expected outcome (−1), and most unfavorable outcome (−2; Kiresuk & Sherman, 1968).

The FWPE was used to assess whether a change in skill competence occurred during the Level II fieldwork experience. The FWPE was developed by a task force commissioned by the AOTA Committee on Education and showed a high correlation ($r = .6$) between Rasch ability measures (Andrich, 1988; Masters & Keeves, 1999) and fieldwork educators’ perceptions of entry-level competence (Atler, 2003). In addition, the FWPE demonstrated good reliability for item calibration and an acceptable goodness of fit (Atler, 2003). The FWPE uses a 4-point Likert scale on which a score of 4 represents the performance level of the top 5% of students that the fieldwork educator has supervised, a score of 3 represents entry-level competence, a score of 2 represents a level of performance that needs improvement, and a score of 1 represents unsatisfactory performance (Atler, 2003). A student is assessed using this scale at midterm and at the end of the fieldwork (Atler, 2003).

Procedure

The University of Utah Institutional Review Board approved the study on February 20, 2014. Consent from 30 students was obtained by the researcher (Jeanette Koski) before setting and scaling the GAS. A cover letter was sent to the 30 fieldwork educators of those students along with the FWPE and the GAS, and consent was implied if the fieldwork educator completed the GAS for the student. A cover letter accompanied an email that was sent to fieldwork educators after the initial data collection to gather information about fieldwork educator demographics.

Occupational therapy students were assigned to fieldwork sites on the basis of their practice preferences, their learning needs, program policies, the features of the sites including the style of the fieldwork educator, and availability of sites. Goals were collaboratively established by the student and the first researcher (Jeanette Koski) to reduce measurement error (King et al., 1999; Malec, 1999). For student goals
to be individualized and still be relevant to the site, each student selected 10 goals from a list of site-specific goals obtained from the AOTA website (AOTA, n.d.). The first researcher scaled 5 of these student-selected goals (King et al., 1999) by determining what behavior would represent each level of the scale; a sample of the scaling is included in Figure 1. The goal levels were reviewed by the second researcher (Lorie Gage Richards) and adjusted until agreement on goal-level definitions was reached.

An explicit assumption made by the research team was that because this Level II fieldwork was the first nonclassroom competency, each student participant had essentially the same baseline: “unable to perform the goals without assistance.” Errors when using GAS may occur if there is no reference point for the final score—in other words, the baseline performance level must be determined to show the magnitude of change (Cytrynbaum, Ginath, Birdwell, & Brandt, 1979).

Fieldwork educators rated the students on the GAS at 3 wk, 1 day, and 3 wk, 3 days, to establish test–retest reliability for that measure. They also completed FWPE and GAS ratings for their students at midterm (6 wk) and at the final (12 wk). Fieldwork educators also completed a survey of demographic information.

Data Analysis

Data were entered into a statistical computer program (IBM SPSS Statistics Version 22; IBM Corporation, Armonk, NY) for analysis. Test–retest reliability for GAS was determined using intraclass coefficient $1, k$ (Gliner, Morgan, & Leech, 2009, p. 160) for each goal for each student. The FWPE items that most closely matched the GAS goals were selected for each student. The mean scores at 6 wk and at 12 wk across these items were computed for each student.

Change in scores on these FWPE items was assessed with a one-way repeated-measures analysis of variance (ANOVA). The mean score for GAS items at 6 wk and at 12 wk for each student was computed. Sensitivity to change on the GAS was tested with a one-way repeated-measures ANOVA.

Results

Twenty fieldwork educators agreed to participate by providing GAS scores (67% response rate). Thirteen of the 20 fieldwork educators who supervised the Level II occupational therapy students filled out the demographic survey. The fieldwork educators had supervised between 1 and 32 Level II students over the course of their time as an occupational therapy practitioner (mean = 7), and they had an average of 13.65 total yr of occupational therapy practice and an average of 7.27 yr of occupational therapy practice at their current location.

Test–retest reliability demonstrated a nearly perfect correlation for the GAS for each of the five goals, showing that each goal was reliable with itself 3 days later (Goal 1, $r = 1.000$; Goal 2, $r = .957$; Goal 3, $r = .953$; Goal 4, $r = .969$; Goal 5, $r = .749, p < .001$). There was a significant amount of change on the FWPE, $F(1, 19) = 97.41, p < .001$. There was also a significant amount of change on the GAS, $F(1, 19) = 36.64, p < .001$. For both measures, scores improved from the 6-wk to the 12-wk measurement period (FWPE, mean change $= 0.90$, standard deviation $= 0.41$; GAS, mean change $= 0.52$, standard deviation $= 0.37$).

Discussion

It is clear that entry-level OTD programs will need to develop a method of measuring whether students meet “individualized specific objectives” (AOTA, 2012, p. 64) on the experiential part of the curriculum. The FWPE evaluation, the current standard for measuring occupational therapy nonclassroom educational experiences, is not suitable for measuring experiential student performance with the potential degree of variability in student goals. Therefore, developing a reliable and valid method of evaluating experiential student performance is critical.

The high correlation between 3-wk, 1-day, and 3-wk, 3-day, scores on the GAS in this short interval during Level II occupational therapy student fieldwork suggests that GAS was a reliable measure for use with occupational therapy students completing this type of nonclassroom experience. The correlation for Goal 5 was likely lower because of greater variability in these goals between students. In addition, though the goals were not specifically weighted, Goal 5 seemed to capture more complex skills than Goals 1–4. Although the correlation was lower for Goal 5, the score remained at a significant level for reliability.

These results indicate that GAS has the potential to be a reliable measure for other types of nonclassroom experiences such as the experiential component of an OTD program. This study’s analysis also shows that GAS was able to identify significant change in a sample in which it was established that change had occurred using the valid FWPE (Atler, 2003). Therefore, GAS may be a sensitive measure of a student’s progress over time and has the potential to be used in evaluating student goal attainment in a nonclassroom environment, that is, the OTD experiential component. However, clear utility of GAS for OTD student evaluation will have to be

![Figure 1. Sample Goal Attainment Scaling goal. Note: FW = fieldwork.](http://ajot.aota.org/)
established directly for the OTD experiential component because it is a different type of nonclassroom experience than Level II fieldwork.

Limitations

Additional research with larger samples will be necessary to validate and generalize this study’s findings. Moreover, the GAS goals could have been more relevant to the site if each site had set its own site goals instead of the researcher using generic site-specific goals suggested by contributors to the AOTA website.

Another limitation was that the time between testing sessions used for test-retest reliability was relatively short, and it is possible that raters remembered previous scores. However, Level II fieldwork is most often a period of rapid skill development for students, and the researchers believed that a longer period between test and retest assessment would have been confounded by actual student learning. It was unfortunate that demographic data were not returned from all the participating fieldwork educators. This limitation made it difficult to determine whether characteristics of the fieldwork educators moderated the effects.

Further study would be beneficial to determine whether psychometric properties of the measure depend on fieldwork educator characteristics. Last, if the GAS is to be used to measure occupational therapy student experiential performance, then a pilot study will need to be performed to determine its validity for that application.

Implications for Occupational Therapy Education

The results of this study have the following implications for occupational therapy education:

- GAS demonstrates good reliability and is sensitive to change when used to measure student goal achievement in nonclassroom experiences.
- Further research is warranted to determine whether GAS is a valid measure of occupational therapy student goal achievement on the entry-level OTD experiential component.

Conclusion

Despite the challenges in setting and scaling goals using GAS, this tool was shown to be a reliable and sensitive measure of occupational therapy student goal attainment on Level II fieldwork, a current nonclassroom occupational therapy educational experience. Fieldwork educators felt that the GAS allowed them to score a student who was performing at the expected level but also to represent a student who was very high functioning. The entry-level OTD experiential component represents another form of nonclassroom experience in which a student will need to set and achieve highly individualized goals. Further work on validating the use of GAS for nonclassroom competency will need to be undertaken within the OTD experiential component itself, but the results of this study indicate the potential use of GAS to measure goal achievement within that component. ▲

References


