Hammill Institute on Disabilities

Career Development and Transition for Exceptional Individuals
1–12
© Hammill Institute on Disabilities 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/2165143417691021
cdtei.sagepub.com
SAGE

Factor Structure and Basic Psychometric Properties of the *Transition Assessment and Goal Generator*

Maeghan N. Hennessey, PhD[1], Robert Terry, PhD[1], James E. Martin, PhD[1], Amber E. McConnell, PhD[1], and Donna M. Willis, BS[1]

Abstract

We examined the theoretical factor structure fit and psychometric properties of the *Transition Assessment and Goal Generator* (TAGG). In the first study, 349 transition-aged students with disabilities, their special educators, and family members completed TAGG assessments, and using exploratory factor analysis (EFA)/confirmatory factor analysis (CFA), a simple eight-factor structure emerged. In 2 subsequent years, independent samples of students with disabilities (Study II, *N* = 257; Study III, *N* = 846), their special educators, and family members completed the TAGG, and the TAGG structure replicated in both years across all versions using CFAs. The results provide evidence of construct validity and reliability. We discuss implications of the results and identify future research needs.

Keywords

transition assessment, assessment development, transition education, annual transition goals

The Individuals With Disabilities Education Act (IDEA; 2004) requires Individualized Education Program (IEP) team members to prepare students with disabilities for postsecondary employment, further education, and, as needed, independent living (Flexer & Baer, 2013). This law established several transition education provisions, including the following three requirements: First, educators must use transition assessments to identify students' interests, strengths, and needs (Rowe, Mazzotti, Hirano, & Alverson, 2015). Second, transition assessment results must be used to develop employment, further education, and, as needed, independent living postsecondary goals authentically representing students' desired postsecondary life (Neubert & Leconte, 2013). Third, IEPs of transition-aged youth must include annual transition goals to address students' transition needs (IDEA Regulations 20 U.S.C. § 1416[a][3][B]; Peterson et al., 2013).

Several issues, however, limit the effective use of transition assessment results to facilitate writing transition sections of students' IEPs. First, the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014) state, "[V]alidity is the most fundamental consideration in developing tests and evaluating tests" (p. 11). As such, ample validity evidence to support the use of assessment results for educational planning is required. Investigation of the internal structure of a transition assessment is the necessary first step to ensure that "a test can indicate the degree to which relations among test items and test components conform to the construct on which the proposed test score interpretations are based" (p. 16). This internal structure should be generalizable to all individuals who may take the assessment in the future, and replication of internal structure findings is needed to verify results and to maintain scientific integrity (AERA et al., 2014; Burman, Reed, & Alm, 2010; Cook, 2014).

Second, federal district court decisions suggest that schools include formal transition assessments, which are assessments with supporting validity evidence (Neubert & Leconte, 2013), as part of the overall transition assessment battery educators complete (Prince, Plotner, & Yell, 2014). Unfortunately, in her nationwide study of most commonly used transition assessments, Martin (2013) found that many were not developed to assess transition-aged students with disabilities, and often little to no validity evidence existed to support using the assessment results in transition planning.

Third, the common practice of identifying students' transition needs in relation to specific postsecondary goals may

[1]The University of Oklahoma, Norman, USA

Corresponding Author:
Maeghan N. Hennessey, The University of Oklahoma, 820 Van Vleet Oval, Room 310, Norman, OK 73019, USA.
Email: maeghan@ou.edu

not result in the most useful annual transition goals. Because students' postsecondary goals often change and actual postschool outcomes seldom match high school goals (Steele, Konrad, & Test, 2005), educators need to consider an assessment approach identifying generalizable behaviors that when learned will assist students to attain multiple postsecondary goals.

A number of researchers (Mazzotti et al., 2015; McConnell et al., 2013; Test et al., 2009) have recently reviewed the literature to identify the generalizable behaviors necessary for annual transition goals that lead to postsecondary employment and further education. Although commonalities exist between the findings in these studies (e.g., a primary emphasis on nonacademic behaviors), McConnell et al.'s (2013) construct list differed from those of Test et al. (2009) and Mazzotti et al. (2015) in two important ways: First, McConnell et al. included studies with designs other than correlational research, expanding the set of possible teachable behaviors that could be included in a transition assessment. Second, definitional constructs and lists of associated student behaviors found in those studies to improve postschool outcomes were developed. A comparison of these skills and experiences with assessments currently used to develop annual transition goals found that no assessment addresses these behaviors (McConnell et al., 2013). Combining this with the dearth of validity evidence supporting the use of commonly used transition assessments suggests a new assessment be created.

Fourth, The Council for Exceptional Children's Division on Career Development and Transition (DCDT) position paper on transition assessment (Neubert & Leconte, 2013) called for multiple stakeholders (e.g., education professionals, students, or family members) to provide information about student performance. Triangulation of assessment results will provide a more accurate representation of transition-aged students with disabilities' interests, strengths, and needs to develop quality IEP transition sections (Simmons, Bauder, & Flexer, 2013).

Transition Assessment and Goal Generator (TAGG) Development

To overcome the four limitations facing users of transition assessments, we compiled assessment specifications, undertook an iterative development process to write, refine, and test the items and develop the scoring methods, and decided upon the administration instructions per the guidelines delineated in the *Standards of Educational and Psychological Testing* (AERA et al., 2014) to develop the online *Transition Assessment and Goal Generator* (TAGG) Professional (TAGG-P), Family (TAGG-F), and Student (TAGG-S) versions. To examine the extent to which the internal structure of the three TAGG versions aligns with behaviors associated with postschool outcomes found in the extant literature, items were developed to assess each of the constructs detailed by McConnell et al. (2013).

We developed the Professional TAGG (TAGG-P) version first because it would use the most complex language and concepts. Items were organized by constructs, with construct definitions preceding the items. Sixty-six of the 75 items used a 5-point Likert-type scale, and instructions asked professionals to rate students' behaviors over the last year using a score of 1, meaning *the student rarely performed the action*, to 5, meaning *the student often performed the behavior or was successful at completing the action*. Nine items required a yes or no answer (e.g., "The student attained at least one transition goal"). Following the recommendations of Clark and Watson (1995), we initially developed an item pool approximately twice as large as we believed necessary.

The Family TAGG (TAGG-F) version used the same items included in the final TAGG-P. Wording for each item was changed to reflect the respondent's role, but the wording remained as close to that used in the TAGG-P as possible. For the TAGG-F, each item began with the stem "My child (followed by a verb) . . ." For example, an item used to assess *Knowledge of Strengths and Limitations* in the TAGG-P was, "The student told someone what he or she does well." This item was modified for the TAGG-F to read, "My child told someone what he or she does well." The TAGG-F used a 5-point Likert-type scale for 66 items, and required a yes or no response for nine items. Instructions to think about actions the child performed in the past year, as well as item anchors, were identical to those provided in the TAGG-P.

Concurrent with the TAGG-F development, we developed the TAGG-S using the same items included in the TAGG-P, with wording revised to be student friendly, written in first person, and written using simpler language. The item used for the TAGG-S paralleling the example above was, "I told someone about the things I do well in school." The final 75-item TAGG-S had a 4.2-grade Flesch-Kincaid reading level. Instructions asked students to think about what they had done or learned over the past year to respond to items by using a 3-point scale, corresponding to "rarely," "sometimes," and "never." Nine items required a yes or no answer.

The purpose of this study was to examine the TAGG's internal structure to determine the relations among items and constructs. We first assessed the factor structure of the three TAGG versions to determine item fit and basic TAGG psychometric properties to provide construct validity evidence. Second, a year later, we replicated the initial factor study with an independent sample collected to confirm the TAGG's factor structure and the score reliability produced from its use by high school students with disabilities, their special educators, and family members. Third, a year after the second replication, we again replicated the initial factor

study with a third independent sample. Below, we present study procedures, describe participants, instrumentation, fidelity checks, and results for each study separately, followed by an overall discussion section.

Method

General Study Procedures

Educator duties. Participating special educators completed nine tasks: (a) watched a 15-min video describing duties, (b) obtained principals' signed agreement for educators and students to participate, (c) facilitated parental consent for family members and their student to participate, (d) completed an educator demographic form, (e) completed student demographic forms, (f) completed the TAGG-P on each student, (g) administered the TAGG-S to each student and provided any needed accommodations or support, (h) facilitated completion of the TAGG-F, and (i) mailed completed materials to us. Professionals received US$30 for each completed packet of assessment materials, which took approximately an hour of their time. Family members and students received a US$10 gift card for their participation. In the first study year, a subsample of respondents participated a second time approximately 3 months after the first TAGG administration, using the same procedures to determine test–retest reliability.

Recruitment. We recruited transition education professionals via email from lists of participants who attended transition education conferences and workshops in 42 states. The recruitment email contained a web link to a 15-min video explaining the research project, their roles, responsibilities, administration instructions, and honorarium for participating in the study. Educators who agreed to participate contacted parents of their students to invite involvement and gain parental consent and student assent or consent.

Completion of demographic forms. Participating professionals completed an educator demographic form asking identifying information, highest education level, position, certifications, and other similar questions. Participating professionals also completed a student demographic form for each of their students involved in the study. The professionals provided information, including students' gender, age, grade, ethnicity or race, participation in general education, disability, and provided a copy of the students' transcripts. Participating educators facilitated completion of a parent demographic form that asked family members to identify their relationship to the student and provide basic demographic information.

Administration of the student TAGG. Participating professionals administered the TAGG-S individually or in groups to the students involved in the study, and provided needed accommodations and support. Before students began answering TAGG items, educators read an instructional script to explain the purpose of the assessment and the directions for completion. Students generally completed the assessment in less than 15 min. During administration, educators encouraged students to think before responding to each item, ask questions, and complete unanswered items if they chose. As needed, educators read the TAGG-S aloud.

Administration of the family TAGG. Each family received a cover letter explaining the purpose of the TAGG-F along with instructions. Family members also received instructions to reflect on their child's behavior and experiences over the last year as they answered each item. Most family members completed the TAGG-F at home, and others completed it at school.

Study I

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) suggest that, after construct development, item creation, scoring, and other basic assessment development steps, the assessment's factor structure should be examined to determine the dimensionality of the construct space and item fit and overall model fit within the construct space. Because the TAGG is a new instrument, psychometric properties also need to be determined.

Study I Method

Participants. A total of 349 transition-aged students with various disabilities from 33 high schools across seven states comprised the student sample for this first study. We collected family and professional parallel measurements on these 349 students from 39 high school special educators and 271 parents. A subsample of these respondents (102 students, 12 educators, and 92 parents) completed the appropriate TAGG versions a second time approximately 3 months after the first administration. See Table 1 for detailed participant demographics. Very few respondents (n = 9) refrained from answering TAGG questions, with usually only one or two missing responses. These incomplete cases were listwise deleted from the analysis, so that only complete cases were analyzed.

Fidelity of study administration and data entry. Educators completed an administration checklist indicating whether each step in the administration process was followed. Overall, teachers reported 98.8% TAGG-P administration fidelity (range = 86%–100%) and 97.7% administration fidelity for the TAGG-S (range = 86%–100%). A member of the research team observed 10 participating educators at 10 different high schools administering TAGG study

Table 1. Educator, Family Member, and Student Demographic Information by Study.

	Educators			Family			Student		
Variable	S1	S2	S3	S1	S2	S3	S1	S2	S3
Sample size	39 *12*	34	100	271 *92*	229	346	349 *102*	257	846
Average age	47 (10.2) *48 (8.8)*	46 (8.7)	45 (10.8)	45 (8.6) *45 (8.1)*	43 (11.3)	44 (10.0)	17 (1.4) *17.5 (1.6)*	16 (3.1)	17.9 (1.0)
% female	94.9 *100*	91.2	94.0				46.4 *47.6*	43.7	38.2
Family respondent									
% mother/stepmother				80 *71*	78	75.5			
% gather/stepfather				11 *16.3*	11	14.8			
Racial/ethnic categories									
% Caucasian	76.9 *75*	85.3	90.0	68 *82.6*	75.1	61.5	67 *73.5*	70.3	70.3
% African American	12.8 *8.3*	11.8	8.0	10 *6.5*	8.3	8.7	17.5 *5.9*	11.4	19.6
% Hispanic	5.0 *16.7*	2.9	1.0	6 *3.3*	5.2	14.6	12 *14.7*	11.4	14.7
% American Indian	2.6 *0*	8.8	3.0	3 *0*	10.9	8.2	4 *0*	13.4	8.6
Student information									
% 9th grade							12 *7.8*	21	1.3
% 10th grade							26 *18.7*	16.3	2.5
% 11th grade							27 *22.5*	25.7	33.8
% 12th grade							35 *52*	35.6	52.4
% 18–21 program							0 *0*	0	9.2
% LD							61 *62*	56.6	52.2
% ID							12 *9.8*	13.1	13.5
% OHI							10 *4.9*	15.2	14.6
% ED							5 *4.9*	6.7	5.3
% other disability							12 *3.9*	8.4	9.5

Note. Other disabilities included are autism, hearing, visual, speech, and traumatic brain injury. Italicized numbers in S1 represent demographic information for subsample participants. S1 = Study I; S2 = Study II; S3 = Study III; LD = Specific Learning Disability; ID = Intellectual Disability; OHI = Other Health Impairment; ED = Emotional Disturbance.

materials with an interobserver agreement of 97% (range = 86%–100%).

To ensure accurate data entry, two researchers independently entered and checked the entire dataset using original data sheets. The percentage agreement between the two researchers was 99.6, and disagreements were resolved using a consensus decision-making process.

Study I Results

Structural analysis of the TAGG-P. We began by submitting the TAGG-P (initial 75-item version) to an exploratory factor analysis (EFA). The initial EFA on the TAGG-P was conducted using maximum-likelihood (EFA-ML) estimation and allowed the factors to correlate using the Promax

rotation using the PROC FACTOR program in SAS®. We investigated factor solutions ranging from 1 to 10 factors, and then compared the values of the Akaike Information Criterion (AIC; Akaike, 1974) and Schwarz's Bayesian Information Criterion (BIC; G. E. Schwarz, 1978) for each of the 10 solutions. Preacher, Zhang, Kim, and Mels (2013) discussed the use of the AIC and BIC in determining the number of factors, suggesting that the AIC tends to find more replicable solutions, whereas the BIC tends to identify the approximately "correct" number of factors. Because we intend to cross-validate our solution across multiple sources and multiple independent samples, we chose to emphasize the AIC, while also using the BIC to help choose among acceptable AIC solutions. The appropriate number of factors using this method (EFA-ML) was determined by noting when both the AIC and the BIC appeared to stop changing dramatically as the number of factors increased, similar to examining a scree plot when using a principal components analysis. As a secondary criterion, factors showing less than three salient loadings (>.25) beyond which the AIC and BIC began to flatten indicated an overfitting of the number of dimensions, and all dimensions beyond that point were discarded.

For the TAGG-P, the EFA-ML analysis suggested that eight factors were the best fit to the data. Subsequent examination of the factor reference structure matrix suggested that 20 items did not substantially load on any of the eight factors, including all items designed to assess *Actions Related to Strengths and Limitations*, and *Utilization of Resources*. Of the remaining 55 items that did load on at least one factor, most items loaded substantially on only one of the eight factors, suggesting a possible simple-structure solution.

To further refine the TAGG-P scale, we submitted the reduced 55-item scale to a new analysis using linear confirmatory factor analysis (CFA) techniques. We examined the overall fit of the CFA models using three indices: root mean square error of approximation (RMSEA) index, for which Steiger and Lind (1980) and MacCallum, Browne, and Sugawara (1996) suggested that values below .08 indicate an acceptable fit and values at or near .05 indicate an excellent fit; Bentler's comparative fit index (CFI; Bentler, 1990), based on the value of the noncentrality parameter, for which values greater than or equal to .90 indicate a good fit; and the Tucker–Lewis index (TLI; Tucker & Lewis, 1973), which compares the fit of the model to an independence (or no factor) model, and for which values greater than or equal to .90 indicate a good fit. Examining the combination of these three fit indices assesses different aspects of model fit. If all indices suggest an acceptable fit, we could be confident that the model adequately reproduces the empirical data.

Examination of the asymptotically standardized residuals (ASRs), obtained after fitting the model, indicated the existence of doublet factors (Landis, Edwards, & Cortina, 2009; Mulaik, 2009) occurring when a pair of items, usually loading on the same factor, share residual or specific variance that cannot be accounted for by the common factor models. In essence, doublet factors typically arise when two items are understood by the participants to represent the same concept, and hence are answered in a highly correlated manner. Two solutions for doublet factors are usually proposed (Landis et al., 2009). The first solution is to fit a model with correlated residuals, which increases the fit of the factor model and adds complexity to that model. The second solution is to eliminate one of the offending items from the doublet pair, keeping the simple structure of the model and reducing the length of the scale. Because reducing the length of the scale has its own virtues (such as simplifying the task for respondents) and dropping items would allow us keep our imposed simple factor structure, we decided to use the latter strategy and deleted items from doublet pairs, which is also appropriate for the population of students for which this assessment was designed. We chose to keep the item written to assess the conceptually advanced behavior in the doublet pair because it encompassed the less advanced behavior. For instance, in one doublet pair, Item 1 was "to develop a plan," with Item 2 being "implement the plan." We kept "implement the plan" as it was the more advanced behavior, and implementing assumes that the plan had been developed.

As a result, an additional 21 items were deleted from the scale, resulting in a 34-item TAGG-P. Once the model was reduced to 34 items, a final CFA of the TAGG-P was investigated. Using MPlus, we estimated factor solutions ranging from one to eight factors using both categorical factor analysis with weighted-least-squares means- and variance-adjusted (WLSMV) estimation and the linear factor model using maximum-likelihood (ML) estimation (to obtain the AIC and BIC statistics, which are not defined for the WLSMV estimator). Each of the eight factors represents a theoretically identified construct in the literature (McConnell et al., 2013). This evidence, coupled with the consistency found between the model–data fit and theoretically defined constructs, suggested that an eight-factor solution was most appropriate. The fit statistics for this linear CFA simple-structure model were acceptable (χ^2 = 1,043.08, *df* = 499; RMSEA = .058, CFI = .92, TLI = .91; see Note 1). Because of the loss of certain items, we revised construct names to better reflect the content of the remaining items: *Knowledge of Strengths and Limitations* became *Strengths and Limitations, Proactive Involvement* became *Interacting With Others, Self-Advocacy* became *Student Involvement in the IEP*, and *Supports* became *Support Community*.

We then refit the model using a categorical factor model, fitting a polychoric correlation matrix and using the robust WLSMV estimator (Asparouhov & Muthén, 2010; Muthén

& Muthén, 2010), because this procedure has been shown to give better estimates of factor loadings as well as less biased fit statistics for categorical data. The fit of the eight-factor simple-structure model to the reduced 34-item scale was acceptable (χ^2 = 1,058.01, *df* = 499; RMSEA = .058, CFI = .974, TLI = .971; see Note 2). Each item loaded on only one factor, and the R^2 values for each item were strong for the most part, indicating that the eight-factor simple-structure solution explained the item variability quite well. The only problematic item with the categorical CFA was Item 24 ("The student expresses wanting a job that matches his or her career interests and skills"), which had a small (e = −.05) nonsignificant negative residual variance within one standard error of 0 and was thus constrained to 0 in the final analysis, which is a standard solution to the problem.

Structural analysis of the TAGG-F. Given our twin goals of factor replicability across sources and the desire to have comparable instruments across informants, we decided to first examine the fit of the final factor model for the TAGG-P version to the TAGG-F version. This analytic approach has at least two positive aspects: First, if the factor model developed for the TAGG-P fit the data approximately as well for the TAGG-F, it would provide evidence for convergent validity of the factor constructs, as the model was empirically built using TAGG-P responses only; that is, it would provide for a partial independent validation of the factor model as the sources of information are partially independent, even though the participants being assessed are the same. If the TAGG-P model fit the TAGG-F responses, it would provide one source of independent verification of its validity. Second, if the TAGG-F factor model held, it would allow us to simplify the assessment process considerably, having the identical number of items and factors for each source of data.

Using the categorical CFA procedure with the WLSMV estimator, the fit of the eight-factor simple-structure model to the reduced 34-item TAGG-F was acceptable (χ^2 = 738.73, *df* = 499; RMSEA = .046, CFI = .977, TLI = .975). Comparing these results with those obtained from the factor model fit to the TAGG-P indicates an extremely similar fit in terms of overall model fit to the data, with all fit indices showing similar or better values. All items loaded positively on only one factor (by design), and no estimation problems occurred.

Structural analysis of the TAGG-S. For the same reasons given previously when analyzing the TAGG-F, we decided to first examine the fit of the final factor model of the TAGG-P to the TAGG-S using the categorical CFA procedure with WLSMV with robust estimation. The fit of the eight-factor simple-structure model to the reduced 34-item student scale was also acceptable, with two caveats. First, Factor 1 (*Strengths and Limitations*) and Factor 8 (*Support Community*) correlated perfectly in the student sample, which we believe resulted because students who could accurately identify their limitations also could accurately identify from whom to seek support. Second, Item 24 again resulted in a small, nonsignificant negative residual variance that was constrained to 0. The fit of the model was acceptable (χ^2 = 787.91, *df* = 499; RMSEA = .045, CFI = .942, TLI = .935). Comparing these results with those obtained from the factor model fit to both the TAGG-P and TAGG-F indicated similar overall model fit to the data. Visual inspection of the factor loadings and R^2 values showed a similar pattern of results to those obtained for the TAGG-P and TAGG-F versions, although the results suggested less salience to the factor loadings and more measurement error in each item.

Reliability: Internal consistency. We assessed the internal consistency of the overall scale score and each subscale score for each version of the TAGG and the 34 items across the eight factors of the Professional and Family versions and the seven factors of the Student version using Cronbach's coefficient alpha (see Table 2). The results showed that the overall scores for each version of the TAGG exhibit internal consistency, with alphas ranging from .89 to .95. Likewise, the reliability estimates of the eight subscale scores for both the TAGG-P and TAGG-F were also acceptable, ranging from α = .60 to α = .93. The exception, however, was with the *Interacting With Others* scale for the TAGG-F (α = .52). Examination of the reliability estimates for the student sample subscale scores suggested lower reliabilities in general, ranging from α = .44 to α = .82; yet the overall alpha level is at a respectable .89 level.

Reliability: Test–retest. A test–retest measure of stability of total TAGG scores was investigated for a subsample of students across the three assessment versions during the first school year in which data were collected (see Table 1 for demographic data). The two administrations averaged 13.7 weeks apart and yielded statistically significant ($p < .01$) and large correlations of .80, .70, and .70 for the TAGG-P, TAGG-F, and TAGG-S, respectively (r = .10 [small], r = .30 [medium], and r = .50 [large]).

Reliability: Agreement across TAGG versions. The total scores across the three TAGG versions showed statistically significant ($p < .01$) correlations. Pearson's product–moment correlation coefficients for Professional–Family (n = 269), Professional–Student (n = 339), and Family–Student (n = 268) TAGG versions yielded the values of .38, .37, and .31, respectively.

Study II

One year later, we conducted a replication study (Study II) to determine whether the factor structure and psychometric properties of the TAGG found in Study I could be replicated

Table 2. Reliability Coefficients for All Three Studies.

	TAGG-P			TAGG-F			TAGG-S		
Factors	S1	S2	S3	S1	S2	S3	S1	S2	S3
Strengths and Limitations	.85	.84	.99	.81	.75	.97	.64	.55	.98
Disability Awareness	.81	.76	.95	.80	.77	.91	.64	.57	.95
Persistence	.93	.94	.99	.91	.90	.98	.82	.82	.99
Interacting With Others	.77	.71	.93	.52	.52	.91	.44	.45	.94
Goal Setting and Attainment	.90	.91	.98	.88	.87	.96	.73	.64	.99
Employment	.74	.72	.81	.62	.50	.78	.56	.55	.91
Involvement in the IEP	.85	.87	.94	.85	.85	.93	.81	.75	.95
Support Community	.68	.64	.94	.60	.62	.87	—	—	—
Overall	.95	.94	.99	.94	.93	.98	.89	.85	.99

Note. Cronbach's coefficient alpha internal consistency rating is generally viewed as follows: $\alpha \geq .9$ = *excellent*, .89–.8 = *good*, .79–.7 = *acceptable*, .69–.6 = *questionable*, .59–.5 = *poor*, and .49 and below = *unacceptable*. Items 25 for TAGG-P and 26 for TAGG-S are not included in this analysis. The factors of Strengths and Limitations and Support Community are combined in the TAGG-S data. TAGG = *Transition Assessment and Goal Generator*; TAGG-P = Professional TAGG; TAGG-F = Family TAGG; TAGG-S = Student TAGG; S1 = Study I; S2 = Study II; S3 = Study III; IEP = Individualized Education Program.

with a new sample of secondary-aged students with disabilities. In Study II, we attempted a direct replication and followed sampling methods and other study procedures as consistently as was possible. Ultimately, the sample size obtained was smaller than that in Study I, and the student respondents were approximately 1 year younger than those in Study I.

Study II Method

Participants and instrument. A total of 257 transition-aged students with various disabilities from seven states comprised the student sample for this study. Family and professional parallel measurements on these 257 students were collected from 34 high school special educators and 229 parents. See Table 1 for additional demographic data. The six incomplete cases in the dataset were again listwise deleted, with 47 students repeating participation.

Fidelity of study administration and data entry. Teachers reported implementing 92.3% (range = 82%–100%) of the administration instructions and procedures for the TAGG-S. Observations of 10 educators were also completed. Educators implemented 90% (range = 71%–100%) of the administration instructions during the observed administrations. Two researchers independently entered 31% (107 cases) of the dataset, with a percentage agreement of 99.8. Disagreements were resolved using a consensus decision-making process.

Study II Results

Replication of the TAGG-P factor structure. When fitting the same factor model found in Study I using categorical CFA with WLSMV robust estimation in the Mplus software package, we found the structure of the TAGG-P in Study I to be an acceptable fit for the data collected from the second sample (χ^2 = 1,236.91, *df* = 499, RMSEA = .067, CFI = .969, TLI = .966), suggesting that the underlying factor pattern of the data remained approximately the same for the two samples. Once again, Item 24 resulted in a small (−.02), statistically insignificant, negative residual variance.

Replication of the TAGG-F factor structure. Results of fitting the factor model from Study I, using categorical CFA with WLSMV robust estimation, to data collected from family members produced an acceptable fit (χ^2 = 742.81, *df* = 499, RMSEA = .052, CFI = .970, TLI = .967) with no modifications to the model, suggesting that the underlying factor pattern structure of the data remained the same across different samples. There were no estimation issues, and all factor loadings were positive.

Replication of the TAGG-S factor structure. Fitting the Study I model using categorical CFA with WLSMV robust estimation of the TAGG-S produced an acceptable fit to the Study II data (χ^2 = 885.34, *df* = 499, RMSEA = .054, CFI = .897, TLI = .884), again suggesting that the factor pattern remained similar across the two samples. Unlike in Study I, Factors 1 and 8 were not perfectly correlated. Item 24 again resulted in a small, statistically insignificant (−.001), negative residual variance.

Reliability. An investigation of the internal consistency of the overall scale score and each subscale score for each version of the TAGG using data collected in Study II produced acceptable results, ranging from α = .86 to α = .93 (see Table 2). Likewise, the reliability estimates of the eight subscale scores for the TAGG-F and seven for the TAGG-S were also acceptable, ranging from α = .62 to α = .94, with

two exceptions. The first exception was with the *Interacting With Others* scale for the family sample (α = .51), and the second was the *Employment* scale (α = .52). These results were similar to those from Study I, with reliability estimates for the student subscale scores lower than those from the professional or family samples, ranging from α = .39 to α = .82. However, the overall alpha level of the TAGG-S was .86, which represents an acceptable reliability level.

Study III

Study III Method

Participants and instrument. As part of a larger investigation, Study III is the second replication and took place 2 years after Study I. We designed Study III to confirm the structure of the TAGG-P, TAGG-F, and TAGG-S versions using categorical CFA techniques with the new sample. Two differences were made to sample selection: (a) we attempted to recruit only student participants in 11th and 12th grades, and (b) we recruited a larger number of participants. A total of 846 transition-aged students with various disabilities across 27 states comprised the student sample for this study. Family and professional parallel measurements on these 846 students were collected from 100 high school special educators and 346 parents (see Table 1 for demographics). As in the previous two studies, the number of incomplete cases was exceptionally small, and those participants were listwise deleted from the dataset.

Study III Results

Replication of the TAGG-P factor structure. The factor structure of the TAGG-P previously found in Studies I and II resulted in an acceptable fit to the data collected from the third sample (χ^2 = 1,190.41, *df* = 499, RMSEA = .065, CFI = .973, TLI = .969). This result suggested that the underlying factor pattern found in Studies I and II remained the same across the third sample. In this sample, Item 24 no longer showed a negative residual variance, most likely due to the larger sample size. Factor loadings and R^2 values for each item are provided in Table 3, and the matrix of correlations among TAGG-P factors in Study III is provided in Table 4. Maintaining these results in a third independent sample gives strong evidence for TAGG-P construct validity.

Replication of the TAGG-F factor structure. The fit of the factor model from Studies I and II to data collected from family members in Study III using categorical CFA with WLSMV robust estimation produced acceptable fit (χ^2 = 914.26, *df* = 499, RMSEA = .060, CFI = .953, TLI = .947). Factor loadings and R^2 values for each item are provided in Table 3, and the matrix of correlations among TAGG-F factors in the third sample is provided in Table 4.

Replication of the TAGG-S factor structure. Again, categorical CFA using WLSMV robust estimation examined the fit of the model to data collected on the TAGG-S in Study III and showed acceptable fit (χ^2 = 870.84, *df* = 499, RMSEA = .040, CFI = .947, TLI = .940). As in all other analyses across samples and informants in Study III, model–data fit was again acceptable, with the addition of no estimation problems for Item 24 in the current data. Table 3 shows the factor loadings and R^2 values for each item, and Table 4 gives the matrix of correlations among the TAGG-S factors.

Reliability. As was done in Studies I and II, we again assessed the internal consistency of the overall scale score and each subscale score for each version of the TAGG using data collected in Study III from the 34 items across eight factors of the Professional and Family versions and the seven factors of the Student version using Cronbach's coefficient alpha (see Table 2). The results show that scores obtained from the overall scales for each version of the TAGG were highly reliable, ranging from α = .98 to α = .99, most likely due to the large number of participants. Likewise, the reliability estimates of the eight subscale scores for both the TAGG-P and TAGG-F were also acceptable, ranging from α = .78 to α = .99. Scores obtained from the TAGG-S subscales were in the same range (α = .91–.99) as those obtained from the TAGG-P and TAGG-F, with an overall alpha of .99.

Discussion

Results indicate that the factor structure of the TAGG largely matched the theoretical framework on which it was designed. With minor differences, we replicated the factor structure across three samples and three sources of information (i.e., parents, educators, and students), providing evidence for factorial and convergent validity. Psychometric analyses indicate that scores exhibit acceptable internal and test–retest reliability. Thus, we are confident that the behaviors included in the TAGG adequately measure the behaviors theoretically identified as important for the postsecondary education and employment of students with disabilities as originally described in McConnell et al.'s (2013) constructs. Moreover, we were able to meet our twin goals of factor replicability across sources and the desire to have comparable instruments across informants. Next, we discuss implications of these findings, suggest possible limitations, and offer future research ideas.

Most transition assessments commonly used to write transition sections of IEPs (prior to the development of the TAGG) have little to no validity evidence to support their use for this purpose (Martin, 2013). Moreover, Prince et al. (2014) identified federal district court cases in which judges questioned the use of only informal transition assessments in secondary transition planning. Conclusions based on the

Table 3. Factor Loadings and R^2 Values for Three TAGG Versions for Study III.

		TAGG-P		TAGG-F		TAGG-S	
Factor	Item	Loading	R^2	Loading	R^2	Loading	R^2
Factor 1	Item 1	.76	.58	.87	.75	.76	.58
	Item 2	.73	.54	.74	.55	.48	.23
	Item 3	.91	.82	.84	.71	.64	.41
	Item 4	.80	.64	.63	.40	.34	.12
Factor 2	Item 5	.73	.53	.86	.75	.82	.68
	Item 6	.82	.67	.91	.84	.60	.36
	Item 7	.68	.47	.68	.46	.67	.45
	Item 8	.48	.23	.58	.34	.42	.18
Factor 3	Item 9	.86	.75	.80	.63	.77	.59
	Item 10	.94	.88	.88	.78	.83	.69
	Item 11	.92	.84	.89	.80	.77	.59
	Item 12	.93	.86	.92	.84	.75	.56
	Item 13	.91	.82	.81	.66	.72	.51
Factor 4	Item 14	.85	.72	.90	.82	.64	.41
	Item 15	.57	.33	.45	.20	.36	.13
	Item 16	.84	.70	.64	.41	.76	.57
Factor 5	Item 17	.86	.73	.84	.71	.77	.59
	Item 18	.94	.89	.81	.65	.61	.37
	Item 19	.96	.91	.87	.75	.80	.65
	Item 20	.94	.88	.82	.67	.62	.39
	Item 21	.91	.82	.89	.79	.79	.62
	Item 22	.57	.32	.66	.44	.25	.06
Factor 6	Item 23	.86	.74	.87	.76	.88	.78
	Item 24	.94	.87	.96	.92	.97	.94
	Item 25	.17	.07	.26	.07	.12	.01
	Item 26	.39	.15	.38	.14	.15	.02
Factor 7	Item 27	.89	.78	.83	.70	.89	.79
	Item 28	.91	.83	.88	.76	.85	.72
	Item 29	.94	.89	.90	.82	.85	.73
	Item 30	.63	.40	.52	.27	.36	.13
Factor 8	Item 31	.77	.59	.75	.56	.75	.55
	Item 32	.65	.43	.60	.36	.65	.43
	Item 33	.78	.61	.69	.47	.69	.48
	Item 34	.40	.16	.55	.30	.24	.06

Note. Factor 1 = Strengths and Limitations; Factor 2 = Disability Awareness; Factor 3 = Persistence; Factor 4 = Interacting With Others; Factor 5 = Goal Setting and Attainment; Factor 6 = Employment; Factor 7 = Student Involvement in the IEP; Factor 8 = Support Community. Factor 8 is with Factor 1 for TAGG-S. TAGG = Transition Assessment and Goal Generator; TAGG-P = Professional TAGG; TAGG-F = Family TAGG; TAGG-S = Student TAGG; IEP = Individualized Education Program.

analyses of these court cases led Prince et al. (2014) to conclude that transition plans "should not be solely dependent on informal measures of students' interests and abilities" but rather "include formal assessments" of student knowledge and skills (p. 46). The results of this study strongly suggest that the TAGG exhibits an internal structure generalizable across multiple samples and informants that can be used to meet Prince et al.'s (2014) recommendation of using formal rather than only informal assessments for transition planning to complete IEP transition sections. Doing so places the use of assessment data collected from a formal assessment in a central position in the IEP transition planning process (Lowenthal & Bassett, 2012).

A second implementation recommendation made by Prince et al. (2014) was to ensure that "student participation in the IEP process is maximized" (p. 46). By including a psychometrically sound student version of the TAGG in addition to professional and family versions, the student's voice can be better heard. The inclusion of a psychometrically sound student version of the TAGG also adheres to DCDT's position that multiple, triangulated results help provide a more accurate representation of

Table 4. Factor Correlations for Three TAGG Versions for Study III.

	Version	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Factor 2	TAGG-P	.88	—					
	TAGG-F	.71	—					
	TAGG-S	.72	—					
Factor 3	TAGG-P	.66	.67	—				
	TAGG-F	.71	.49	—				
	TAGG-S	.81	.44	—				
Factor 4	TAGG-P	.62	.65	.73	—			
	TAGG-F	.53	.39	.68	—			
	TAGG-S	.75	.52	.79	—			
Factor 5	TAGG-P	.72	.76	.81	.72	—		
	TAGG-F	.75	.62	.86	.65	—		
	TAGG-S	.80	.58	.86	.68	—		
Factor 6	TAGG-P	.58	.57	.51	.59	.61	—	
	TAGG-F	.55	.45	.53	.52	.67	—	
	TAGG-S	.72	.42	.75	.63	.75	—	
Factor 7	TAGG-P	.49	.58	.47	.46	.61	.45	—
	TAGG-F	.56	.53	.52	.47	.64	.72	—
	TAGG-S	.47	.53	.55	.46	.69	.47	—
Factor 8	TAGG-P	.72	.72	.85	.81	.76	.60	.54
	TAGG-F	.65	.66	.87	.70	.86	.64	.64
	TAGG-S	.90	.61	.73	.63	.77	.77	.60

Note. Factor 1 = Strengths and Limitations; Factor 2 = Disability Awareness; Factor 3 = Persistence; Factor 4 = Interacting With Others; Factor 5 = Goal Setting and Attainment; Factor 6 = Employment; Factor 7 = Student Involvement in the IEP; Factor 8 = Support Community. Factor 8 is with Factor 1 for TAGG-S. TAGG = Transition Assessment and Goal Generator; TAGG-P = Professional TAGG; TAGG-F = Family TAGG; TAGG-S = Student TAGG; IEP = Individualized Education Program.

students' interests, strengths, and needs when developing IEP documents (Neubert & Leconte, 2013). Including all three versions allows a shared vision of the student's future to be discussed and planned (Simmons et al., 2013).

Finally, the results of our studies provide strong evidence that the internal structure of the three versions of the TAGG adequately represents the behaviors it is intended to measure (McConnell et al., 2013). Specifically, results from two replications of the initial analysis show strong evidence that the structure of the three versions of the TAGG is generalizable to individuals who are not in these samples (Burman et al., 2010). The fact that the characteristics of the three samples are not identical provides further evidence that this assessment can be used for a variety of students who have disabilities (N. Schwarz & Strack, 2014).

Limitations and Future Research

Because the development of psychometrically sound instruments is an iterative and ongoing process, where assessments are continually analyzed and revised as long as the assessment is in use (AERA et al., 2014), there will always be a need to collect and analyze more data to refine the assessment and demonstrate evidence of validity. This ongoing, iterative process necessitates each study undertaken in the development of an assessment has limitations suggesting the need for future research. In this study, for example, limitations of the characteristics of our samples did not allow for the investigation of the extent to which subgroups with different demographic characteristics (e.g., student disability category or racial/ethnic background) will respond similarly to TAGG items. Future research should also examine the development of growth in skills assessed by the TAGG across time. The significant correlations across TAGG versions suggest that the parents, professionals, and students evaluate students somewhat similarly given their different perspectives and experiences with the students, which future research needs to investigate. Finally, although the results of this study lead us to be confident in the internal structure of this assessment, the extent to which scores on the three versions of this instrument predict that students' future postsecondary education and employment remain unknown. To address this limitation, longitudinal data on student outcomes after high school graduation are required to determine whether TAGG construct scores predict further education and employment.

Despite these limitations, we believe that the results of this study met our purpose and contribute to the assessment of the internal structure of the three TAGG versions. All three versions exhibit similar internal structures, and the

factor structures found in the first sample were substantially replicated twice. Because of the stable nature of the underlying constructs assessed by the TAGG, educators can feel confident in using the results of this assessment to guide writing annual IEP transition goals for their students.

Authors' Note

The University of Oklahoma owns the *Transition Assessment and Goal Generator* (TAGG) copyright, and OU's Office of Technology Development receives and distributes funds from TAGG sales per Regents and State of Oklahoma policies.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: A grant from the Institute of Education Sciences, National Center for Special Education Research, Grant Award R324A100246, in part, supported this research project.

Notes

1. To fully investigate the structure of the *Transition Assessment and Goal Generator–Professional* (TAGG-P), hierarchical and bifactor models were tested in addition to the simple-structure model. The fit of the eight-factor, simple-structure model was better than that of either the hierarchical (χ^2 = 1,315.14, *df* = 512; root mean square error of approximation [RMSEA] = .067, comparative fit index [CFI] = .89, Tucker–Lewis index [TLI] = .85) or bifactor (χ^2 = 1,186.92, *df* = 494; RMSEA = .064, CFI = .91, TLI = .85) models.
2. Hierarchical and bifactor models were again tested using categorical confirmatory factor analysis (CFA) with similar results. Specifically, the fit of the eight-factor, simple-structure model was better than that of either the hierarchical (χ^2 = 1,303.76, *df* = 519; RMSEA = .068, CFI = .96) or bifactor (χ^2 = 1,153.63, *df* = 493; RMSEA = .064, CFI = .97) models. These results suggest that the eight-factor, simple-structure model be used in all subsequent analyses.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Asparouhov, T., & Muthén, B. (2010). *Plausible values for latent variables using Mplus* (Mplus technical report). Retrieved from www.statmodel.com/download/Plausible.pdf

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *107*, 238–246. doi:10.1037/0033-2909.107.2.238

Burman, L. E., Reed, W. R., & Alm, J. (2010). A call for replication studies. *Public Finance Review*, *38*, 787–793.

Clark, L., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309–319. doi:10.1037/1040-3590.7.3.309

Cook, B. G. (2014). A call for examining replication and bias in special education research. *Remedial and Special Education*, *35*, 233–246. doi:10.1177/0741932514528995

Flexer, R. W., & Baer, R. M. (2013). Transition legislation and models. In R. W. Flexer, R. M. Baer, P. Luft, & T. J. Simmons (Eds.), *Transition planning for secondary students with disabilities* (4th ed., pp. 22–45). Boston, MA: Pearson.

Individuals With Disabilities Education Act, 20 U.S.C. § 1400 (2004).

Landis, R. S., Edwards, B. D., & Cortina, J. M. (2009). On the practice of allowing correlated residuals among indicators in Structural Equation Models. In C. E. Lance & R. J. Vanderberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity, and fable in organizational and social sciences* (pp. 195-214). New York, NY: Routledge.

Lowenthal, A., & Bassett, D. (2012). Transition assessment: Using gap analysis to enhance effective transition planning. *Intervention in School and Clinic*, *48*, 30–37. doi:10.1177/1053451212443131

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149. doi:10.1037/1082-989x.1.2.130

Martin, J. D. (2013). *Examining the measurement invariance of the Transition Assessment and Goal Generator across percent of time spent in general education* (Doctoral dissertation). University of Oklahoma, Norman.

Mazzotti, V., Rowe, D., Sinclair, J., Poppen, M., Woods, W., & Shearer, M. (2015). Predictors of post-school success: A systemic review of NLTS2 secondary analyses. *Career Development and Transition for Exceptional Individuals*, *39*, 196–215. doi:10.1177/2165143415588047

McConnell, A. E., Martin, J. E., Juan, C. Y., Hennessey, M. N., Terry, R. A., El-Kazimi, N. A., . . . Willis, D. M. (2013). Developing the Transition Assessment and Goal Generator (TAGG): Building student behavior constructs as the first step in developing a new transition assessment. *Career Development and Transition for Exceptional Individuals*, *36*, 174–187. doi:10.1177/2165143412468147

Mulaik, S. A. (2009). *The foundations of factor analysis*. New York, NY: CRC Press.

Muthén, L. K., & Muthén, B. (2010). *Mplus user's guide: Version 6*. Los Angeles, CA: Author.

Neubert, D. A., & Leconte, P. J. (2013). Age-appropriate transition assessment: The position of the Division on Career Development and Transition. *Career Development and Transition for Exceptional Individuals*, *36*, 72–83. doi:10.1177/2165143413487768

Peterson, L. Y., Burden, J. P., Sedaghat, J. M., Gothberg, J. E., Kohler, P. D., & Coyle, J. L. (2013). Triangulated IEP transition goals: Developing relevant and genuine annual goals. *Teaching Exceptional Children*, *45*(6), 46–57. doi:10.1177/004005991304500606

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*, 28–56. doi:10.1080/00273171.2012.710386

Prince, A. M. T., Plotner, A. J., & Yell, M. L. (2014). Postsecondary transition and the courts: An update. *Journal of Disability Policy Studies*, *25*, 41–47. doi:10.1177/1044207314530469

Rowe, D. A., Mazzotti, V. L., Hirano, K., & Alverson, C. Y. (2015). Assessing transition skills in the 21st century. *Teaching Exceptional Children*, *47*, 301–309. doi:10.1177/0040059915587670

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464. doi:10.1214/aos/1176344136. MR468014

Schwarz, N., & Strack, F. (2014). Does merely going through the same moves make for a "direct" replication? Concepts, contexts, and operationalizations. *Social Psychology*, *45*, 305–306.

Simmons, T., Bauder, D., & Flexer, R. W. (2013). Collaborative transition services. In R. W. Flexer, R. M. Baer, P. Luft, & T. J. Simmons (Eds.), *Transition planning for secondary students with disabilities* (4th ed., pp. 176–198). Boston, MA: Pearson.

Steele, R. B., Konrad, M., & Test, D. W. (2005). An evaluation of IEP transition components and post-school outcomes in two states. *The Journal for Vocational Education Special Needs Personnel*, *27*(2), 4–18.

Steiger, J. H., & Lind, J. M. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, *Iowa City, IA*.

Test, D., Mazzotti, V., Mustian, A., Fowler, C., Kortering, L., & Kohler, P. (2009). Evidence-based secondary transition predictors for improving postschool outcomes for students with disabilities. *Career Development and Transition for Exceptional Individuals*, *32*, 160–181. doi:10.1177/0885728809346960

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1–10. doi:10.1007/BF02291170