Psychological Test and Assessment Modeling, Volume 63, 2021 (3), 305-336

Latent dirichlet analysis to aid equity in the identification for Gifted Education

W. Holmes Finch[1], Maria E. Hernández[1] Finch, Brian F. French[2], Claire Braun[1]

Abstract

The identification of students for participation in gifted education programs has traditionally involved the use of some combination of cognitive ability measures and teacher recommendations. Research has demonstrated that these approaches may limit the participation of students from diverse ethnic and socioeconomic backgrounds. In order to address this issue, efforts have been made to develop assessments and techniques that are more equitable for all students. The current study was focused on the development of a scale based on parental descriptions of aspects giftedness displayed by their child. Topic-modeling via latent dirichlet analysis was used to extract themes from 13 written responses by parents. The topics were then used in a nonlinear principal components analysis to develop a scale score. Evidence for validity was then investigated for these scores in a variety of ways. Results demonstrated that scores differed between student groups as expected, and correlations with other measures associated with giftedness were in the expected direction and magnitude. Implications and utility of this scale are discussed.

Keywords: Gifted identification, Topic modeling, Validity assessment, Nonlinear principal components analysis, HOPE scale

[1] Ball State University, USA

[2] Washington State University, USA

The identification of children for participation in gifted education (GE) programs has traditionally relied on measures of cognitive ability (e.g., standardized assessment scores), teacher nominations, or both sources of information. The use of these approaches may contribute to disproportional underrepresentation of some populations due to issues with the assessments, as well as to long standing inequities in the educational system (Peters et al., 2012; Dale, et al., 2014). Therefore, in an effort to make gifted identification more equitable for the population as a whole, educators are beginning to favor identification methods that represent the multifaceted nature of giftedness (Subotnik, et al., 2011). Several scales have been suggested as potentially useful for this purpose, including the HOPE (Having Opportunities Promotes Excellence) Scale, which was designed to measure a wide array of gifted attributes in a culturally sensitive and inclusive manner (Peters & Gentry, 2010). There is evidence that the HOPE scale score may be more culturally sensitive than the traditional standardized test scores that are frequently used in GE identification (Peters & Gentry, 2010). In addition to rating scale items that are used to calculate the total HOPE score, the instrument also includes a set of open-ended items that allow parents to provide short descriptions of their children's gifted behaviors/skills. The purpose of this study was to develop a scale score based upon these qualitative descriptions, and then to assess the validity evidence for its use as an additional tool in the identification of children for GE programs.

Assessment of giftedness

A key component of GE is the identification of those who should participate in such programs. For many years, identification of children for GE was accomplished by evaluation of performance on standardized measures of intelligence (Terman, 1926; Hollingworth, 1942; Cattell, 1987). In its simplest form, such identification practice involved individuals being given a test of cognitive ability, and then identified as gifted if their score on the assessment exceeded a predetermined cut-value. Subsequently, the definition of what constitutes giftedness has broadened, with the emergence of various multi-faceted definitions of giftedness coming to the fore (Marland, 1972). Such work expanded the definition of giftedness to include a wider array of facets beyond the traditional intellectual sphere, including the arts, creative endeavors, and interpersonal interactions (Plucker & Callahan, 2014). Subotnik, Olszewski-Kubilius, and Worrell (2009) further describe the construct of giftedness as being a multifaceted combination of both cognitive and psychosocial variables (e.g., culture, language, socioeconomic status).

Despite these advances in the field of gifted identification, there is not a consistent approach that is applied across educational institutions and organizations (McBee at al., 2012; NAGC, 2015). For this reason, schools and districts in the United States apply a wide variety of approaches to identifying students for participation in GE programs. In some cases, these identification approaches reflect relatively narrow definitions of giftedness, relying primarily on standardized assessments of cognitive

ability, and/or teacher recommendations. In turn, this dependence on such traditional approaches may disproportionately under-identify diverse (e.g., lower SES, racial/ethnic minorities, ELL) students (Ford et al., 2008). The use of these identification methods tends to yield homogenous populations for participation in GE programs, whereas more inclusive techniques could lead to a more representative population of students who qualify for participation in GE programs (McBee et al., 2012; Scott et al., 2014).

The issue of appropriate and timely identification of children for giftedness is particularly important given that when gifted diverse students are identified early and then placed into GE programs, their academic outcomes are better than when identification occurs later (Hanushek & Rivkin, 2006; Hansen & Toso, 2007; Janus & Offord, 2007; Merriman, 2012; Ford, 2014). In addition, identification of children for participation in GE programs when they are young leads to more diverse students being identified, as well as to better educational outcomes for these individuals (Horowitz, 1987, Hanushek & Rivkin, 2006; Subotnik et al., 2011; Merriman, 2012. In other words, early identification (e.g., preschool age) for participation in GE programs appears to be beneficial both in terms of increasing the number of diverse students in those programs, and with respect their long-term educational outcomes.

Traditional processes for GE identification occurs in two steps: (1) Nomination of individual students by their teachers (NAGC, 2013; McBee et al. 2016), and (2) Testing of students using standardized assessments. These traditional methods for identifying gifted individuals pose a number of potential problems, particularly for diverse students. In particular, it has been found that teachers may systematically under refer Black or Latinx students (Ford et al., 2008), and students from low socioeconomic backgrounds (McBee, 2006). The use of standardized assessments for identifying gifted students is also not without problems. Specifically, such measures often do not have representative norming samples, thus calling into question the applicability of their normed scores for use with diverse students (Peters & Gentry, 2010; Peters & Engerrand, 2016). In addition, many such tests have a relatively high linguistic load, which may penalize students from more diverse backgrounds (Hadaway & Marek-Schroer, 1992; Ford et al., 2008). These limitations have led some to call for the use of nonverbal measures of cognitive ability (Naglieri & Ford, 2003; Shaunessy et al.,2004; Flanagan & Ortiz, 2013; Naglieri & Ford, 2015; Cao et al., 2017). Researchers have found that the use of assessments such as the *Wechsler Intelligence Scale for Children-Revised*, and the Differential Ability Scales (DAS-II) for GE identification results in more equitable results for diverse students (Keith et al., 1999; Edwards & Oakland, 2006; Keith et al., 2010; Baron & Leonberger, 2012; Dale et al., 2014). Finally, because conceptions of giftedness have expanded beyond mental ability to include characteristics such as motivation, creativity, artistic talent, etc. (Renzulli et al. 2009), culturally sensitive rating scales might also function to compensate for the limitations of traditional ability tests by supplementing information regarding these other facets of giftedness.

Given the limitations associated with the use of teacher nomination in conjunction with standardized assessments for GE identification, some researchers have suggested that additional useful information for this purpose may be obtained from parents. In

particular, it has been suggested that parents may be able to speak to aspects of their child's giftedness that are not easily measured by a standardized test, or that a teacher working with a classroom of students may not have an opportunity to observe (Chan, 2000). Parents may also be able to provide educators with an additional source or rich, insightful information about a child's potential giftedness that could be used in the identification process.

Despite the potential utility of information obtained from parents, many commonly used rating scales designed to obtain feedback from educators may not be applicable for parents in rating the giftedness of their young children (Pfeiffer & Petscher, 2008; Pfeiffer et al., 2007), particularly those from diverse backgrounds (Ford, 2013). Given these limitations, the HOPE scale was developed as a tool for use with diverse populations (Peters & Gentry, 2010). The original measure consisted of a 12-item teacher-rating scale for parents of school-aged children. It was then adapted for use with preschool children, through use of open-format responses to supplement the original rating scale items for this study. The purpose of these open-ended questions was to obtain information from parents about aspects of giftedness that their child possesses, and that may not be readily measureable using standardized assessments of cognitive ability and academic achievement. In turn, by allowing parents to reflect on aspects of giftedness that their children might manifest, and which are not captured in traditional measures and approaches for GE identification, it was anticipated that such scales might be particularly useful for culturally diverse children. Scores on the HOPE scale itself do demonstrate evidence of validity and utility for identifying students who qualify for GE programs (Peters & Gentry, 2010).

The goal of the current study was to extend upon this earlier work through the development of a scale score based upon responses to the open ended HOPE scale items. Such a score may reflect aspects of giftedness that parents are in a particularly good position to see, and which may not be reflected in other assessments used for gifted identification. In addition, the score could provide unique information about giftedness that is more culturally sensitive than what is obtained by traditional assessments or teacher recommendations. It is anticipated that the result is an additional tool that educators can use in the identification of children for participation in gifted programs, and one that is particularly well suited to use with preschool age children from diverse backgrounds. The development of this scale was based upon the use of topic modeling, which is discussed briefly below.

Topic Modeling

Topic modeling (TM) falls within the broader statistical family of text mining. It is based in the broader fields of machine learning and natural language processing, and is designed to provide insights into underlying patterns of word use within text. TM is used with a set of individual documents, which are collectively known as a corpus. Underlying TM is an assumption that the distribution of words across the documents

in the corpus is at least in part a function of a set of common topics (Hofmann, 1999). Researchers use TM to identify these topics based on how words are used together, and then topics are characterized based upon the words most commonly associated with them. TM yields estimates of the probability of words being associated with each topic (β), and the probability of each document being associated with each topic (γ). Documents are assigned to the topic for which the value of γ is largest.

TM parameters can be estimated using a variety of approaches, with perhaps the most common being latent dirichlet analysis (LDA). LDA is based in the Bayesian framework, and provides a flexible parameter estimation paradigm that is particularly useful for estimating relatively complex models (e.g., more topics), and for use with smaller samples (Blei et al., 2003) as is the case in this study.

Study Goals

The primary goal of this study was to develop, using TM in conjunction with nonlinear principal components analysis, a scale (HOPEtm) for use in GE identification based upon an open-ended parental descriptions of their children's gifted attributes. This scale would ideally provide an additional piece of information for educators to use in identifying children for participation in GE programs. It would differ from methods that are typically used for this purpose (e.g., cognitive ability tests, teacher recommendations) by incorporating parents' observations of their child's behavior in a variety of contexts that prior research has found to be associated with giftedness. The scale would therefore provide unique and potentially culturally sensitive information that is not available from standardized tests, or even teacher recommendations. The second goal of this study was to examine validity evidence for scores obtained from this open-ended scale by assessing (a) HOPEtm score relationships with popular assessments used in GE identification, (b) HOPEtm score mean differences for students identified and not identified for GE programs, and (c) relationships of the HOPEtm scores with variables that should not determine participation in GE programs, including ethnicity, gender, parental education, language spoken in the home, and family income. Taken together, this evidence would support an extrapolation inference in a validity argument framework.

Method

Participants

Parents of children between the ages of 4 years, 0 months to 5 years, 11 months were included in the study. A cohort of 178 subjects in a small Midwestern city were obtained through a GE screening program in a local school district. A majority (n = 153) of participants were self-referred through a screening process at a local public

laboratory school (Site A), with the remaining participants ($n = 25$) coming from local public preschools in more racially and culturally diverse areas. Interested families completed requisite forms (i.e., informed consent, release of information, HOPE Scale, and demographic form) on site, and testing was completed with these children during their school day.

Table 1 contains demographic information about the sample. The majority of the children included in the study were white/Caucasian, and spoke English in the home. In addition, the majority of mothers (65.2%) had a Bachelor's or Graduate degree, whereas 53.9% of the fathers had this level of education. With respect to siblings, 45.9% of the subjects were only children, 39.6% had one sibling, 8.1% had two siblings, and 6.4% of the children had three or more siblings. The mean income for families included in the study was $72,625.64, and the median was $65,000.

Table 1:
Descriptive Statistics for Sample

Group	Percent
Ethnicity	
White/Caucasian	72.5%
Other Ethnicity	27.5%
Home language	
English	86.0%
Other	24.0%
Highest level of maternal education	
<High school	3.9%
High school	8.9%
Some college	12.4%
Associate's degree	7.9%
Bachelor's degree	26.4%
Graduate degree	38.8%
Not reported	1.7%

Highest level of paternal education	
<High school	4.5%
High school	11.2%
Some college	16.3%
Associate's degree	9.0%
Bachelor's degree	25.3%
Graduate degree	28.6%
Not reported	5.1%
Number of siblings	
None	45.9%
One	39.6%
Two	8.1%
Three or more	6.4%

Variable	Mean	Standard Deviation
Income	$72,625.64	$53,849.06
HOPE	43.3	10.7
WPPSI-IV FSIQ	106.0	14.8
DAS-II GCA	109.3	13.3

Measures

Several measures were administered to the participating children and their parents, including the *Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition* (WPPSI-IV) the *Differential Ability Scales - Second Edition* (DAS-II), and the HOPE scale. The WPPSI-IV was developed to measure cognitive abilities in children ages 2 years, 6 months to 7 years, 7 months (Wechsler, 2012), and yields a measure of overall intelligence (*g*), as well as measures of the individual cognitive abilities that comprise the overall intelligence quotient. The WPPSI-IV standardization sample included identified gifted individuals (Syeda & Climie, 2014), and thus it is considered appropriate for use in identifying gifted students. Prior research has revealed that the

WPPSI-IV has exhibited evidence of both validity and reliability (e.g., Syeda & Climie, 2014; Wechsler, 2012).

In addition to the WPPSI-IV, The *Differential Ability Scales, Second Edition* (DAS-II) was administered to participants in this study. It was originally developed to measure cognitive abilities in children ages 2 years, 6 months through 17 years, 11 months (Elliot, 2007b). The version of the DAS-II used in this study consisted of 6 subtests comprising General Cognitive Ability (GCA), Verbal Ability (Gc), Nonverbal Reasoning (Gf), and Spatial Ability (Gv). Study participants were administered the Verbal Comprehension, Picture Similarities, Naming Vocabulary, Pattern Construction, Matrices, Copying, Early Number Concepts, Matching Letter-Like Forms, and Phonological Processing subtests. These subscales were then used to calculate the GCA and School Readiness scores. Additionally, the Picture Similarities, Picture Concepts, Matrices, and Copying subtests were used to calculate the Special Nonverbal Composite.

The *HOPE Scale* includes a 13-item rating instrument developed by Peters & Gentry (2010) to identify aspects of academic achievement and social giftedness in students K-5. Peters and Gentry (2010) developed this scale for use by teachers in order to provide an alternative assessment for identifying ethnically and economically diverse gifted students. In the current study, an adapted version of the HOPE scale was used with parents of participating children. Prior work has demonstrated that this version of the scale has relatively high levels of reliability (e.g., Cronbach's α values between 0.8 and 0.9), and a variety of validity evidence, including correlations with subsequent performance in gifted programs, unified factor structure (Peters & Gentry, 2010). In addition to the rating scale items, parents were also asked to provide responses to open ended questions about various aspects of giftedness exhibited by their children. These items asked for specific examples of each component of giftedness that was measured by the HOPE scale. The scale, including both the rating scale items and the open ended items, appears in the appendix to this manuscript.

Finally, study participants were asked to provide demographic information, including their children's gender, race/ethnicity, parent education levels, family income level, languages spoken by the child, number of siblings in the home, and number of years the child attended preschool. These items were completed by the children's parent(s) or guardian(s). They were designed to obtain culturally relevant information for the children and their families participating in the study.

Data analysis

The data analysis used to first develop the HOPEtm scale, and subsequently assess its validity included several steps. First, TM (Oh, et. al., 2017; Finch et al., 2018) via LDA was used to extract themes from each of the 13 open-ended responses. For each item, participants were assigned to the topic for which they had the highest probability. Thirteen different corpuses (corresponding to the open-ended items) were used

in this study, each consisting of the parent/guardian responses to the open-ended items.

Prior to fitting the TM using LDA, the data were processed, using standard procedures (Srivastava & Sahami, 2009). First, punctuation and digits were removed from the corpus. In addition, common articles used in English were also removed, including "the", "and", "are", "a", "by", "for", "with", "our", "that", "there", "this", and "them". In addition to removal of specific words, punctuation, and digits, the remaining terms in the text were stemmed, meaning that prefixes and suffixes were removed. Stemming the words ensures that those conveying the same idea are treated as the same word by the LDA algorithm.

Once the data were processed, TMs were fit to each open ended question using LDA, with R software version 3.2.2 through the packages `tm` (Feinerer 2011), `topicmodels` (Grün & Hornick, 2017), and `ldatuning` (Nikita, 2016). The prior distribution for β was dirichlet with α of 50/k, and the prior for γ was dirichlet with δ of 0.1 (Griffiths and Steyvers, 2004). Convergence of the MCMC estimates was assessed through the use of a trace plot for each of the model parameters, and convergence was obtained for each. The LDA was fit with two chains, each with a total of 15,000 replications, and a burnin of 5,000 samples. The remaining 10,000 samples were thinned such that every 10^{th} value was kept, leaving 1,000 samples for estimating the posterior distribution of each model parameter.

The optimal number of topics for each item was identified in part using the density based statistic of Cao et al., (2009) and perplexity. In addition to the use of statistical tools, a content review of the terms associated with each topic was also conducted in order to ensure that the results were conceptually meaningful This content review involved an examination of the theoretical and conceptual consistency of the word groupings associated with each topic. Each family's response to each open-ended item was classified as belonging to the topic for which it had the highest probability. In order to characterize the topics, the probabilities of each word being generated by the individual topics (β) were calculated, as was the frequency of each topic in the sample (γ). Once the words that were most strongly associated with each topic were identified, we examined the actual texts (i.e., open ended item responses) belonging to each topic in order to ascertain how these words were used in practice by the respondents.

The topics that came from the TMs described above were then treated as nominal variables in a nonlinear principal components analysis (NLPCA; Mori et al., 2016), to identify a small set of components that captured the variance/covariance in the set of nominal topics. The `homals` function in the `homals` R library was used to fit the NLPCA. Several solutions were fit to the data, differing by the number of components retained. For each solution, the eigenvalues and proportion of variance explained in the observed data were obtained. The number of retained components was determined by the proportion of variance explained by the components. Because methods that are preferred for determining the number of latent dimensions in factor analysis, such as parallel analysis, have not been investigated for use with NLPCA, we elected to

use the approach of relying on the proportion of variance in the observed variables explained by the latent trait(s). A topic set was presumed to be associated with a component if the loading was 0.32 or higher, meaning that the component accounted for at least 10% of the variance in the topic set (Tabachnick & Fidell, 2013).

In order to investigate associations with other variables validity evidence, component scores from the NLPCA were correlated with scores from the WPPSI-IV, the DAS-II, and the Likert item portion of the HOPE scale. It was anticipated that scores on the HOPEtm would be moderately positively correlated with these scores, and more strongly positively correlated with the HOPE scale score. Component means were also compared between children who were identified as gifted using traditional criteria employed by their school. Finally, relationships between the components and gender, ethnicity, parental education, and family income were examined using analysis of variance (ANOVA), and correlation coefficient, respectively. It is expected that the correlation between the HOPEtm score will exhibit little to no correlation with demographic factors, and that scale means were not significantly different between genders, ethnicities, and parent education groups.

Results

Topic modeling

The first step in the construction of the HOPEtm scale was the identification of topics for each of the open-ended items. Table 2 displays the topics that were identified for each of the items, along with the 3 most common terms associated with these topics, and the percent of the sample whose responses were in the topic. Based on these words, the topics were given general identifiers, which appear in the first column of the table. In subsequent analyses, these topics were treated as nominal item responses. Finally, each person was assigned to the topic for which they had the highest probability based on the topic model. From these results, it appears that for the majority of the open-ended HOPE items, the distribution of individuals into the topic was relatively equal across the sample.

Table 2:

Topics with 3 most common words for each open ended HOPE item, and percent of sample in the topic

Topic	3 most common terms	Percent of sample
Q1. Two examples showing potential for remarkably high academic performance		
T1 Reading	Read, words, books	23.0%
T2 Vocabulary	Memorize, Question, Vocabulary	19.1%
T3 Puzzles	Puzzle, Word, problems	25.3%
T4 Math	Count, math, numbers	19.7%
T5 Playing games	Name, games, play	12.9%
Q2. Two examples of curiosity and questioning		
T1 Learn words	Words, Learn, Everything	39.3%
T2 Curious over different things	See, Different, Curious	31.5%
T3 Reading	Love, Learn, Read	29.2%
Q3. Two examples when child demonstrates empathy		
T1 Family	Brother, Sister, Sad	22.5%
T2 Crying	Ask, Cry, Tell	23.6%
T3 Friends at school	Friend, School, Help	21.9%
T4 Comfort upset friend	Upset, Comfort, Friend	32.0%
Q4. Two examples when child demonstrates compassion		
T1 Help friends	Want, Help, Friend	26.4%
T2 Hug friends	Hug, Play, Friend	21.3%
T3 Give toys	Ask, Give, Toy	32.0%
T4 Concern for sad people	Concern, People, Sad	20.2%
Q5. Two examples of desire to work with advanced concepts and materials		

T1 Words	Learn, Love, Words	29.2%
T2 Building/creating	Build, Play, Legos	27.0%
T3 Reading	Book, Read, Like	21.3%
T4 Science/puzzles	Question, Science, Puzzle	22.5%
Q6. Two examples of child questioning authority		
T1 Cleaning/playtime	Clean, Play, Rule	16.3%
T2 Questioning teachers and other people	Teacher, Question, People	31.5%
T3 Bedtime/playtime	Bedtime, Ask, Play	17.4%
T4 Questioning parents	Mom, Parent, Question	19.1%
T5 Ask reasons	Told, Reason, Ask	15.7%
Q7. Two examples of eagerness to explore new concepts		
T1 New Interests	New, Recent, Interest	17.4%
T2 Trying	Learn, Love, Try	30.3%
T3 Watching Programs	Watch, Video, Show	10.7%
T4 Building/Creating	Build, Art, Play	15.2%
T5 Using Toys Differently	Work, Different, Toy	16.4%
Q8. Two examples of strong sense of social justice and fairness		
T1 Sharing	Get, Everyone, Share	48.9%
T2 Fairness	Play, Toy, Fair	21.3%
T3 Friends/Family	Friend, Sister, Child	29.8%
Q9. Two examples of using alternative processes		
T1 Playing with puzzles	Puzzle, Use, Play	10.1%
T2 Building/Creating	Make, Build, Toy	18.0%
T3 Coloring/Drawing	Color, Paper, Draw	40.4%
T4 New/Novel	New, Different, Instead	14.6%

T5 Problem Solving	Problem, Solve, Something	16.9%
Q10. Two examples of insightfulness and intuitiveness		
T1 Interactions	Tell, People, Family	19.1%
T2 Understanding new things	Understand, New, Something	26.4%
T3 Creating/Puzzling	Build, Puzzle, Draw	54.5%
Q11. Two examples of "thinking outside the box" (thinking creatively)		
T1 Building/Creating	Use, Build, Make	25.0%
T2 Stories	Creative, Work, Story	32.9%
T3 Playing	Play, Toy, Lego	42.1%
Q12. Two examples of demonstrating intense interests		
T1 Music	Music, Read, Play	20.2%
T2 Building	Love, Build, Lego	17.7%
T3 General Interests	Want, Know, Question	15.1%
T4 Puzzles/Games	Interest, Puzzle, Game	21.4%
T5 Dinosaurs/Animals	Dinosaurs, Book, Animals	25.6%
Q13. Two examples of outstanding talent in school subject		
T1 Math/Arithmetic	Math, Subtract, Count	18.9%
T2 Reading	Read, Word, Letter	25.2%
T3 Art	Art, Color, Draw	14.7%
T4 Reading/Math	Love, Book, Math	22.7%
T5 Writing/Language	Write, Language, Spell	18.5%

Nonlinear Principal Components Analysis

For the NLPCA, models from 1 to 4 components were fit to the HOPE topic data. Results suggested the presence of a single component that explained a total of 69.6% of variance in the nominal topic variables. When up to four components were fit to the data, the second, third, and fourth components explained an additional 7.1%, 6.8%, and 5.6% of the variance in the topic variables. Given these results, it appears that a single component solution is optimal.

The loadings for the single component model appear in Table 3. All of the items, except for number 13 (outstanding talent in school subject), had loadings of 0.32 or higher, meaning that the component accounted for at least 10% of the variance in each. The largest loadings were associated with item 2 (examples of curiosity and questioning), item 9 (use of alternative processes), item 3 (examples of empathy), and item 1 (potential for remarkably high academic performance). The lowest loadings (other than for item 13) belonged to items 7 (eagerness to explore new concepts), 5 (desire to work with advanced concepts/materials), 8 (strong sense of social justice), and 6 (questioning authority). Considering these results, it appears that all of the items, apart perhaps from number 13, contributed to the HOPEtm scale, and that those contributing the most were associated with curiosity, alternative ways of thinking, and empathy. Given the evidence supports a unitary scale for the open ended HOPE items, a HOPEtm component score was calculated for each individual in the study.

Table 3:

Nonlinear Principal Component Loadings for Topic variables associated with Open-ended HOPE Scale Items

Item	Loading
Q1. Two examples showing potential for remarkably high academic performance	0.68
Q2. Two examples of curiosity and questioning	0.94
Q3. Two examples when child demonstrates empathy	0.76
Q4. Two examples when child demonstrates compassion	0.54
Q5. Two examples of desire to work with advanced concepts and materials	0.48

Q6. Two examples of child questioning authority	0.51
Q7. Two examples of eagerness to explore new concepts	0.45
Q8. Two examples of strong sense of social justice and fairness	0.50
Q9. Two examples of using alternative processes	0.86
Q10. Two examples of insightfulness and intuitiveness	0.55
Q11. Two examples of "thinking outside the box" (thinking creatively)	0.57
Q12. Two examples of demonstrating intense interests	0.55
Q13. Two examples of outstanding talent in school subject	0.27

Convergent validity evidence for HOPEtm

In order to assess the evidence available for the convergent validity of the HOPEtm scale, relationships with variables that are considered to be associated with aspects of giftedness were assessed. As described in the Methods section, this evidence consisted of the correlations between scores on the HOPEtm scale obtained by NLPCA and a variety of cognitive ability measures. Correlations between the HOPEtm component score with scores on the WPPSI-IV Full scale IQ and the DAS-II General Cognitive Ability (Table 4) were moderate and positive (Cohen, 1988), suggesting that individuals with higher scores on the topic component performed better on the tests of cognitive ability. These results are in keeping with what was anticipated. Correlations between the topic component and subtests that might be less culturally sensitive (Flanagan & Ortiz, 2013) were also positive and moderate in size (Table 4). Thus, higher scores on the HOPEtm component were associated with higher scores on verbal comprehension, working memory, processing speed, and cognitive proficiency, as expected. Finally, the correlation between the HOPE score obtained using the closed form items and the HOPEtm component score was large, suggesting that these two aspects of the HOPE scale were measuring a similar construct, but not completely redundant.

In addition to developing convergent validity evidence using correlations between the HOPEtm score and a variety of cognitive assessment measures, further evidence was also collected by comparing the mean HOPEtm scores between children who were ultimately identified as gifted, versus those who were not. It is important to note that the decision to identify children for participation in the gifted education program did

not involve the use of the HOPE scale score, nor the open-ended responses that were provided by the parents. Thus, identification of gifted status (yes or no) was completely independent of the HOPEtm component score, making it useful for assessing validity evidence via known group differences. HOPEtm component means between children identified and not identified as gifted through the school's standard screening were significantly different (p=0.004) with a large Cohen's d effect size (Cohen, 1988) of 0.89, as seen in Table 5. The mean HOPEtm score for those children identified independently as being gifted were 0.02 (SD=0.05) as compared to those who were not so identified, -0.01 (SD=0.03).

Table 4:

Correlation Coefficients for Topic Component with WPPSI-IV, DAS-II, and HOPE Scales

Variable	Topic Component
WPPSI-IV Full Scale IQ	0.312
DAS-II General Cognitive Ability	0.224
WPPSI-IV Verbal Comprehension	0.213
WPPSI-IV Working Memory	0.362
WPPSI-IV Processing Speed	0.283
WPPSI-IV Cognitive Proficiency	0.228
HOPE Scale	0.882

Discriminant validity evidence for HOPEtm

Finally, in order to ensure equity in gifted identification, there should not be relationships between construct irrelevant demographic factors (gender, family income, ethnicity) and the GE identification mechanism. In other words, scales used as part of GE identification should not yield different results for children who come from different demographic, socioeconomic, or ethnic backgrounds, as these represent construct irrelevant sources of variation, with respect to giftedness. The correlation

coefficient for the relationship between household income and the HOPEtm scale score was 0.02, indicating that there was not a relationship between the two variables. In contrast, the correlation between family income and FSIQ was 0.267, and for GCA it was 0.170. These correlations suggest that there was a stronger relationship between these traditionally used measures for GE identification and income, than was the case for the HOPEtm score. In addition, the very weak correlations between HOPEtm scores and these demographic factors were as expected.

Table 5 contains the results of the ANOVA model used to compare the means of gender, ethnicity, and mother's education level for each of HOPEtm, HOPE, FSIQ, and GCA. Gender was coded as female/male, ethnicity was coded as white/non-white, and mother's education was coded as less than a college degree, and college degree or higher. The ANOVA results revealed that there were not statistically significant differences between females and males, or white and non-white subjects on any of the measures. However, children whose mothers had a college degree or more, had higher mean scores on both FSIQ and GCA. In contrast, the means on both HOPEtm and the HOPE scale were not statistically significantly different between the two maternal education groups. With respect to effect size, Cohen's *d* for the HOPEtm and HOPE scales were in the negligible range, based on Cohen's (1988) guidelines. On the other hand, *d* was in the small effect range for FSIQ and GCA on both gender and ethnicity, and for FSIQ for maternal education. The value of *d* fell in the moderate range for GCA with respect to mother's education. Therefore, we can conclude that children whose mothers had less than a college degree had significantly lower means on both the FSIQ and GCAM, and that this difference was of a moderate size, based on Cohen's guidelines. Such was not the case for the HOPEtm or HOPE scales, where there were no statistically significant differences between the two maternal education groups.

Table 5:
Mean (standard deviation), and Cohen's d for HOPEtm, HOPE, FSIQ, and GCA scales by selected demographic variables

	HOPEtm	HOPE	FSIQ	GCA
Gender	p=0.85, d=-0.03	p=0.99, d=-0.03	p=0.21, d=-0.23	p=0.42, d=-0.22
Female	-0.03 (1.02)	43.17 (10.38)	104.59 (12.36)	110.25 (11.01)
Male	0.00 (0.99)	43.20 (11.19)	107.95 (16.80)	108.55 (15.58)
Ethnicity	p=0.71, d=-0.07	p=0.84, d=0.04	p=0.34, d=-0.20	p=0.11, d=-0.29
White	-0.03 (1.01)	43.17 (10.03)	106.67 (13.82)	110.26 (12.65)
Non-white	0.04 (1.03)	43.62 (12.29)	103.68 (17.50)	106.35 (15.68)
Mother Ed	p=0.88, d=0.03	p=0.56, d=-0.14	p=0.004, d=-0.44	p=0.003, d=-0.54
<College	0.01 (1.00)	42.13 (13.52)	100.96 (14.27)	104.05 (13.65)
College De-gree	-0.01 (1.02	43.59 (9.94)	107.33 (14.68)	111.06 (12.89)

Discussion

The goal of this study was to develop a scale for assisting educators in determining whether an individual should be placed in a GE program. The scale was based on parental descriptions of their child's behaviors and skills in areas associated with giftedness, collected in the form of 13 short answer items. The development of the scale

involved the use of TM and NLPCA to first develop a set of topics using the open ended item responses, and then to combine them into a single scale. Results of the TM revealed the presence of coherent topics for each item, based upon the open-ended responses provided by parents. The number of topics ranged between 3 and 5 across the 13 items. Secondly, the NLPCA yielded evidence of a unitary dimension underlying the 13 topic sets. This dimension explained approximately 70% of the variance in the item topics, with all but one item having a component loading greater than 0.32. Given this strong evidence of an underlying dimension, a score reflecting children's gifted behaviors, as described by parents, was created using the results of the NPCA. These scores were then investigated with respect to validity evidence to support an extrapolation inference with the scores. Specifically, the HOPEtm scores were correlated with variables that are generally associated with giftedness, including scores on widely used cognitive assessments. These correlations were positive and fell in the moderate range, suggesting that the HOPEtm score was associated with cognitive ability as measured by the WPPSI and DAS-II. Furthermore, the HOPEtm score was strongly related to the HOPE score obtained using the rating scale items, providing further evidence that the HOPEtm was measuring some aspects of the giftedness construct. With respect to group differences, the HOPEtm mean was higher for those who had been identified as gifted using other methods, and was not significantly different for groups for which differences on giftedness would not be expected, including by gender, ethnicity, maternal education, or income. Such was not the case for the WPPSI or DAS-II, both of which were associated with maternal education level and family income.

Taken together, the results summarized above provide support for the use of the HOPEtm scale as an additional piece of information that may be useful for helping educators identify children for GE programs. Indeed, there is some evidence that it may be less susceptible to the influences of diversity with respect to family income and education than is the case with more traditional tools for assessing giftedness in children. In addition, the HOPEtm accommodates a wider definition of what it means for young children to be gifted, without apparently sacrificing sensitivity for identification. Children who were in gifted education programs had higher mean scores on the HOPEtm scale score than did those who were not. Finally, the topics themselves are of use to educators, as they provide information regarding the ways in which a particular child may be gifted, thereby yielding more information to teachers, gifted coordinators, and others about how a particular child exhibits their giftedness.

It is important to state here that we are not proposing the HOPEtm as a replacement for other methods of GE identification. It is unclear whether, by itself, this new scale would be sufficient to make such identification accurately. However, we would suggest that the HOPEtm does provide unique information about giftedness that cannot be captured by measures of cognitive assessment or the rating scale items on the HOPE scale, and that this additional information is associated with being identified for a GE program. Thus, we see the HOPEtm as providing educators with another source of insight about a child's propensity to thrive in a GE setting. This insight would come without some of the construct irrelevant variance around diversity that is

known to impact the more traditional measures in such an evaluation process. In addition, it provides educators with information very specific to an individual, which is not possible with objective scale scores, including both traditional cognitive assessments or the rating scale Hope items. Therefore, teachers, gifted education coordinators, and others who make such decisions may be particularly interested in scores obtained from the HOPEtm scale when decisions around diverse students need to be made.

Directions for future research

This study provides an examination at what has proven to be a promising tool for assisting in GE assessment, using an innovative approach to scoring free response items. However, more work in this area is needed. For example, the scale should be applied to a new sample of children to replicate results and support generalizability. In such a study, parents would be asked to complete the HOPE scale, including the 13 open ended items, their responses would then be applied to the topic model obtained in the current study, and the results used in conjunction with the NPCA model to calculate the HOPEtm scale score. The validity evidence for this score would then need to be investigated much as it has been in the current study. A second avenue for future research would involve replicating this study with a sample taken from a different population. For example, the performance of the HOPEtm scale with somewhat older children, children from more diverse backgrounds, and particularly children from English language learner families all need to be examined, to ensure that the results presented here continue to hold. The demographics of our sample revealed that the parents had a high level of education and income compared to typical families in many school districts. This is a limitation of the current study, and one which should be addressed in future research with the HOPE and HOPEtm scales. Finally, future research should also examine a wider array of validity evidence for the HOPEtm, including its relationships with additional measures of cognitive assessment, as well as with the performance of children in school after their placement in a GE program. This type of evidence would provide further information regarding the utility of the scale in helping educators to determine GE placement.

Conclusions

The determination of placement in a GE program is extremely important for the ultimate success and thriving of gifted children. Prior research suggests that some of the more common tools and methods for doing so remain inequitable for certain subgroups within the broader population (Plucker, 2012). The current study was designed to develop and to investigate a new scale that incorporates parental insights into the process of giftedness assessment. The results presented above demonstrated that this scale, the HOPEtm, shows promise for use as an additional piece of information in GE identification programs. In particular, it may be helpful for use with diverse

populations where standard cognitive assessment measures and teacher recommendations may not be sufficiently accurate. Although more work with this scale certainly needs to be done, it does appear that the HOPEtm may be a useful addition to the gifted educators' toolkit.

References

Baron, I.S. & Leonberger, K.A. (2012). Assessment of intelligence in the preschool period. *Neuropsychology Review, 22*, 334-344. doi: 10.1007/s11065-012-9215-0

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research 3*, 993–1022.

Cao, T.H., Jung, J.Y., & Lee, J. (2017). Assessment in gifted education: A review of the literature from 2005 to 2016. *Journal of Advanced Academics, 28*(3), 163-203. doi: 10.1177/1932202X17714572

Cattell, R.B. (1987). *Intelligence: Its structure, growth, and action.* New York, NY: Elsevier.

Chan, D.W. (2000). Exploring identification procedures of gifted students by teacher ratings: Parent ratings and student self-reports in Hong Jong. *High Ability Studies, 11*(1), 69- 82.

Dale, B.A., Hernández Finch, M., McIntosh, D.E., Rothlisberg, B.A., & Finch, W.H. (2014). Utility of the stanford-binet intelligence scales, fifth edition, with ethnically diverse preschoolers. *Psychology in the Schools, 51*(6), 581-590. doi: 10.1002/pits.21766

Edwards, O.E., & Oakland, T.D. (2006). Factorial invariance of Woodcock-Johnson III scores for African American and Caucasian Americans. *Journal of Psychoeducational Assessment, 24,* 358-366.

Feinerer, I. (2011). tm: Text Mining Package. R package version 0.5-5., URL http://CRAN.R-project.org/package=tm.

Finch, W.H., Hernández Finch, M.E, McIntosh, C.E., & Braun, C. (2018). The use of Topic Modeling with Latent Dirichlet Analysis with Open-Ended Survey Items. Translational Issues in Psychology, 4(4), 403-424.

Flanagan, D.P., Ortiz, S., & Alfonso, V.C. (2013). Essentials of cross-battery assessment: Third edition New York: Wiley.

Ford, D.Y., Grantham, T.C., & Whiting, G.W. (2008). Culturally and linguistically diverse students in gifted education: Recruitment and retention issues. Council for Exceptional Children, 74(3), 289-306.

Griffiths, T.L. & Steyvers, M. (2004). Finding scientific topics. Proceedings of the National Academy of Sciences of the United States of America, 101, 5228-5235.

Grün, B., & Hornik, K. (2017). Topicmodels. R Package version 02.6.,. URL https://cran.r-project.org/web/packages/topicmodels/index.html.

Hadaway, N. & Marek-Schroer, M.R. (1992). Multidimensional assessment of the gifted minority student. *Roeper Review, 15*(2).

Hansen, J.B. & Toso, S.J. (2007). Gifted dropouts: Personality, family, social, and school factors. *Gifted Child Today,* 30(4), 30-41. doi: 10.4219/gct-2007-488

Hanushek, E.A. & Rivkin, S.G. (2006). School quality and the black-white achievement gap. NBER Working Paper 12651.

Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the Twenty-Second Annual International SIGIR Conference*, August, Berkeley, CA.

Hollingworth, L.S. (1942). *Children above 180 IQ.* New York, NY: World Book.

Janus, M. & Offord, D.R. (2007). Development and psychometric properties of the early development instrument (EDI): A measure of children's school readiness*. *Canadian Journal of Behavioral Science, 39*(1), 1-22. doi: 10.1037/cjbs2007001

Keith, T.Z., Low, J.A., Reynolds, M.R., Patel, P.G., & Ridley, K.P. (2010). Higher-order factor structure of the Differential Ability Scales-II: Consistency across ages 4 to 17. *Psychology in the Schools*, *24*(7), 676-697. doi: 10.1002/pits.20498

Keith, T.Z., Quirk, K.J., Schartzer, C., & Elliott, C.D. (1999). Construct bias in the differential ability scales? Confirmatory and hierarchical factor structure across three ethnic groups. *Journal of Psychoeducational Assessment, 17,* 249-268.

McBee, M.T. (2016). What you don't look for, you won't find: A commentary on card and giuliano's examination of universal screening. *Journal of Advanced Academics, 27*(2), 131-138. doi: 10.1177/1932202X16634141

McBee, M.T. (2006). A descriptive analysis of referral resources for gifted identification screening by race and socioeconomic status. *The Journal of Secondary Gifted Education, 17*(2), 103-111.

McBee, M.T., Shaunessy, E., & Matthews, M.S. (2012). Policy matters: An analysis of district-level efforts to increase the identification of underrepresented learners. *Journal of Advanced Academics, 23*(4), 326-344. doi: 10.1177/1932202X12463511

Marland, S.P., Jr. (1972, August). *Education of the gifted and talented-volume 1: Report to the Congress of the United States by the U.S. Commissioner of Education.* Washington, DC: Department of Education, Department of Health, Education, and Welfare. (ERIC Document Reproduction Service No. ED 056243)

Merriman, L. (2012). Developing academic self-efficacy: Strategies to support gifted elementary school students.

Mori, Y., Kurodo, M., & Makino, N. (2016). *Nonlinear Principal Components Analysis and its Applications.* Singapore: Springer.

Naglieri, J.A. & Ford, D.Y. (2015). Misconceptions about the Naglieri Nonverbal Ability Test: A commentary of concerns and disagreements. *Roeper Review, 37,* 234-240. doi: 10.1080/02783193.2015.1077497

Naglieri, J.A. & Ford, D.Y. (2003). Addressing under-representation of gifted minority children using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly, 47,* 155-160.

National Association for Gifted Children. (2015). State of the states in gifted education 2014-2015: Policy and practice data. Retrieved from http://www.nagc.org/sites/default/files/key%20reports/2014-2015%20State%20of%20the%20States%20%28final%29.pdf

Nikita, M. (2016). Ldatuning: Tuning of the Latent Dirichlet Allocation Models Parmaeters. R

Package version 02.0.,.URL https://cran.rproject.org/web/packages/ldatuning/index.html

Oh, J.S., Stewart, A.E., & Phelps, R.E. (2017). Topics in the Journal of Counseling Psychology,

1963-2015. *Journal of Counseling Psychology, 64(6), 604-615.*

Peters, S.J. & Gentry, M. (2010). Multigroup construct validity evidence of the hope scale: Instrumentation to identify low-income elementary students for gifted programs. *Gifted Child Quarterly, 54*(4), 298-313. doi: 10.1177/0016986210378332

Peters, S.J. & Gentry, M. (2012). Group-specific norms and teacher-rating scales: Implications for underrepresentation. *Journal of Advanced Academics, 23*(2), 125-144. doi: 10.1177/1932202X12438717

Peters, S.J. & Engerrand, K.G. (2016). Equity and excellence: Proactive efforts in the identification of underrepresented students for gifted and talented services. *Gifted Child Quarterly, 60*(3), 159-171. doi: 10.1177/0016986216643165

Pfeiffer, S.I. & Petscher, Y. (2008). Identifying young gifted children using the gifted rating scales- preschool/kindergarten form. *Gifted Child Quarterly,* 52(1), 19-29. doi: 10.1177/0016986207311055

Pfeiffer, S.I., Petscher, Y., & Jarosewich, T. (2007). The gifted rating scales—preschool/kindergarten form: An analysis of the standardization sample based on age, gender, and race. *A Journal on Gifted Education, 29*(3). 206-211. doi: 10.1080/02783190709554410

Plucker, J.A. & Callahan, C.M. (2014). Research on giftedness and gifted education: Status of the field and considerations for the future. *Exceptional Children, 80*(4), 390-406. doi: 10.1177/0014402914527244

Renzulli, J.S., Siegle, D., Reis, S.M., Gavin, M.K., & Sytsma Reed, R.E. (2009). An investigation of the reliability and factor structure of four new scales for rating the behavioral characteristics of superior students. *Journal of Advanced Academics, 21*, 84-108.

Scott, M.T., Peters, S.J., & Waterman, C. (2014). Combining scores in multiple-criteria assessment systems: The impact of combination rule. *Gifted Child Quarterly, 58*(1), 69-89. doi: 10.1177/0016986213513794

Shaunessy, E., Karnes, F.A., & Cobb, Y. (2004). Assessing potentially gifted students from lower socioeconomic status with nonverbal measures of intelligence. *Perceptual and Motor Skills, 98,* 1129-1138.

Srivastava, A.N. & Sahami, M. (2009). *Text mining: Classification, clustering, and applications.* Boca Raton, FL: CRC Press Taylor & Francis Group.

Subotnik, R. F., Olszeswski-Kubilius, R., & Worrell, F.C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Intereset, 12,* 3-54. doi: 10.1177/1529100611418056

Syeda, M.M. & Climie, E.A. (2014). Review of Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition. *Journal of Psychoeducational Assessment*, *32*, 265-272. doi:10.1177/0734282913508620

Tabachnick, B.G. & Fidell, L.S. (2013). *Using Multivariate Statistics*. Boston, MA: Pearson.

Terman, L.M. (1926). *Mental and physical traits of a thousand gifted children. Vol. 1. Genetic students of genius* (2nd ed.). Stanford, CA: Stanford University Press.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence—Fourth Edition*. San Antonio, TX: The Psychological Corporation.

Worrell, F.C. (2009). Myth 4: A single test score or indicator tells us all we need to know about giftedness. *Gifted Child Quarterly, 53,* 242-244. doi: 10.1177/0016986209346828