Scale development: Assessment of gifted preschoolers

Caglar Cetinkaya*, Akdeniz University, Faculty of Education Sciences, Konyaalti, Antalya, Turkey
https://orcid.org/0000-0002-1943-7873

Cagdas Cetinkayab, Akdeniz University, Faculty of Education Sciences, Konyaalti, Antalya, Turkey

Suggested Citation:

Received from September 19, 2020; revised from October 22, 2020; accepted from November 16, 2020. Selection and peer review under responsibility of Prof. Dr. Adile Aşkım Kurt, Anadolu University, Turkey, ©2020 Birlesik Dunya Yenilik Arastirma ve Yayincilik Merkezi. All rights reserved.

Abstract

Human intelligence and talent are notions that have been studied in various disciplines like psychology and education overtime. Developing better understanding on these concepts pass through self-exploration of mankind and researchers use fundamental and complex scientific process skills in this process. In this regard, the study aims at developing a measurement tool by which indicators of gifted and talented students can be observed in a valid and reliable way. The research hypothesized four factored giftedness structures and then following technical operations and statistical computations were performed. Item poll was created after an extended literature review and analysis of existing evaluation scales. Items were revised, modified, improved and probable problematic ones were discarded throughout pilot study, and final template of the scale was obtained. Data was collected from 258 preschool/kindergarten children in Antalya province. Findings revealed sufficient and strong evidence for validity and reliability that the measurement tool developed was theoretically supported and a good measure of giftedness.

Keywords: Children; Gifted; Identification; Preschool; Scale development; Talented.
1. Introduction

There has been a great deal of study and debate surrounding education of gifted and talented children who can be described simply as children revealing upper or better performance in certain areas or promising high potential in future (Brody & Stanley, 2005) or not (Borland, 2005). Despite devoted struggles, there is still an ambiguity in who they really are and how they can be distinguished. To recognize them, many conceptualized and put forward their theoretical models which can be categorized according to what their focuses are. In first wave-models (domain-general), giftedness is elucidated at the basis of definition hence first-wave researchers strive to find answers primarily to the question of ‘what is giftedness?’. Being gifted in various ways is mentioned in second wave (domain-specific) models, and third wavers (system) add psychological elements they value to the accumulation of first and second wave researchers’ model ideas and do not deny the importance of mastering on a specific content. Finally, the first three generations of giftedness researchers’ ideas are mostly accepted by fourth wave (developmental) researchers who concentrate primarily on talent development in their giftedness comprehension. Though there seem a great harmony among all models stated above, they are differentiated from each other in several points. No intellective abilities and creativity which accompany conventional intellectual abilities are valued in system and developmental models, whereas knowledge base in single area is acknowledged in domain-specific models. In addition, while being gifted and talented is perceived as an achievement in domain-specific and system models, it can be usually said that developmental researchers regard giftedness as potential, and talent as the product (Kaufman & Sternberg, 2008; Tezer, Özden & Atasoy, 2019).

Identification is the first step for serving gifted and talented children and another living issue elaborated by connoisseurs around the world. Variety and uncertainty in models of giftedness urge experts to develop different sorts of identification methods as well as instruments. Formerly, intelligence tests used to admit children to gifted programs in general (Çetinkaya, 2013). However, it can be easily noticed today that multidimensional assessment tools and methods are replacing traditional or IQ based ones, which is the reflection of evolution in giftedness models. In study of Kuo et al. (2010), for instance, three staged identification procedures in which different techniques such as checklist, interview, observation, portfolio, group and individual intelligence tests was applied to include children to the program. When the study is analyzed, it is seen that identification begins with nomination in which a trait rating scale by which teachers and parents assess children is used. As contemporary identification comprises sequential processes, starting identification with scientifically sound tool would increase accuracy rate of following steps.

Identification of gifted preschoolers is another side of the issue that is like looking for a needle in a haystack (Çoşar, Çetinkaya & Çetinkaya, 2015). Although early intervention in special education is said to be universally supported (Pfeiffer & Petscher, 2008), research cannot be sure whether identification at such a young age is necessary and criticize the issue from various perspectives. Sattler (1988) opposed high reliance on intelligence tests and stability of measured traits in preschool children. Fluctuations in IQ scores of young children is dramatically tremendous comparing with those after especially age 5. Therefore, it is suggested that measures should be taken while IQ scores measured prior to age 5 are being interpreted. According to Perleth, Schatz, and Mönks (2000) IQ scores correlate positively to age of children, but negatively to time interval of measurement. They also propound that researchers also should keep their skepticism on long-term works and even instruments promising high predictive qualities since there is limited number of longitudinal studies in which young, gifted children are screened at least several years to unveil indicators and possible determinants of giftedness and achievement (Şahin & Çetinkaya, 2015; Özcan & Zaroglu, 2021).

Considering what is mentioned above, there are three main reasons that push the researcher for performing this study. It can be said that, firstly, conducting nomination by using scientifically sound instrument is the one more importance should be placed (Keser & Erdem, 2019; Attar, 2019), and more sophisticated works performed on it to increase the success rate of following steps in gifted identification. Secondly, assessment troubles which do not stem from person itself can be addressed
thanks to developing instrument which fits on cultural and social environment since giftedness perception is culturally a dependent fact (Hertzog, 2008; Pfeiffer & Blei, 2008). Thirdly, no similar instrument to GTCEI which was constructed or validated by quantitative methods has been found in Turkey. In this regard, the aim of this study is to develop statistically valid and reliable assessment tool assisting educators at maximum level in identification process and investigate psychometric properties of it.

2. Method

2.1 Participant and Sampling

Population of the study is children studying in preschool institutions in Antalya province of Turkey whose ages ranging from 48 to 82 months. Data was obtained from 258 voluntary preschool students (n= 140 female, n= 118 male) from 11 private educational institutions. Sampling was multi staged in which convenience sampling was used for the main part of the study and stratified nonrandom sampling methods for concurrent validity inspection were performed. In this context, 57 children were reassessed by their teachers for concurrent validity examination of the scale.

The sample of the study, like that of Gifted Rating Scales – Preschool/ Form (GRS-P) developed by Pfeiffer and Jarosewich (2003), was divided into 5 groups and stratified within four 6-month age bands: 48-53 months, 54-59 months, 60-65 months, 66-71 months, and one 72-82 months. It was aimed to make comparisons between product of the study and GRS-P which was also attempted to adapt into Turkish culture twice in different studies by Alma (2015) as doctorate, and Karadağ (2015) as master’s dissertation.

2.2 Instruments

*Gifted and Talented Children Evaluation Instrument (GTCEI)*; Multi-dimensional model of giftedness is adopted in the instrument measuring intellectual ability, academic ability, creativity and artistic talent. The main aim of the scale is to nominate possible gifted and talented children in those areas; however, it is suitable to use as screening tool. It has 41 items in total with 4 subscales which comprise 9, 10, 11, and 11 items orderly. GTCEI has seven-point Likert type format and was labelled according to frequency of observed action as 1: Never, 2: Rarely, 3: Sometimes but infrequently, 4: Neutral, 5: Sometimes, 6: Usually, and 7: Always.

Intellectual ability refers to student’s far-reaching mental skills or capabilities; rated aspects of intelligence in the scale are nonverbal fluid reasoning, verbal knowledge, fluid reasoning, knowledge, quantitative reasoning, visual-spatial processing, and working memory. Academic ability on the other hand refers to student's sufficiency in school-related subject-matter. Developmental domains (cognitive, language, social-emotional, motor, self-care) acknowledged by Turkish Ministry of National Education (MoNE) in National Curriculum for preschool children are rated aspects of the academic ability.

Creativity refers to originality, innovativeness and uniqueness in children's actions, thoughts or productions. Originality, fluency, and detailing in behaviors and abstract thinking are some facets of creativity scale. Artistic talent refers to student's ability in drawing, painting, sculpture, musical plays, and drama.

2.3 Scale development and administration procedure

Foremost concern in producing the instrument was deciding which one is the reasonable? Adaptation of an existing contextually appropriate instrument or developing new one. Although there is rigorous rationale for instrument adaptation and many advantages of it (Humbleton & Patsula, 1999), the researcher came to a decision that developing new tool would be more suitable for following reasons. At first, behavioral expectations in items of other scales do not properly fit Turkish national education curriculum. For example, although ‘invented spelling’ is expected from children in GRS-P, no objective
containing reading or writing skills is included in the Turkish curriculum. Additionally, motivation is not regarded as a type of giftedness but the drive that takes person into success by either the researcher or others like Pfeiffer and Petscher (2008).

Various approaches have been proposed by many researchers for scale development (Clark & Watson, 1995; DeVellis, 2003). One study (Netemeyer, Bearden & Sharma, 2003) suggested four-step scaling procedures which was also selected to follow. It includes construction of definition and content domain in the first, generation and judgment of measurement items in the second, development and refinement of the scale by conducting pilot study in the third, and finalization of the scale by using some confirmatory methods in the final step. In this regard, four factored theoretical structure of giftedness was presumed first. Secondly, item poll was created after extend literature review and analysis of existing evaluation scales. It was aimed to create trait list depends on empirically proven, mostly longitudinal, studies in the widest range, after elimination of items in the pool 51 items were left as a result. The draft scale was reviewed by preschool, gifted, measurement and assessment, and Turkish education Ph.Ds. working in different universities. After no contextual violation was observed by them, it was decided to continue with pilot study in which Exploratory Factor Analysis (EFA) and Item-Test Correlations were calculated. 2 items were removed due to both inadequate factor loadings and negative item-test correlations with the whole scale. During the pilot study, non-familiarity of teachers with giftedness related concepts and complexity of language used in the scale were determined. Therefore, an example of each item has been written and included in the instrument to make behavioral indicators more concrete, easy to understand and teacher friendly. The final version of the scale containing 41 items was obtained. In the main part of the study, 22 teachers from 11 different institutions assessed 258 students by using GTCEI. Items was in mixed order when the scale is given to the teachers to prevent them from rating children monotonously, hence heaping of the items was expected under appropriate scales.

2.4 Data Analysis

SPSS was used for all statistical computations except Confirmatory Factor Analysis f(CFA) and common factor variances in EFA. The following statistical and technical operations were executed to examine validity and reliability of measurement performed by using GTCEI.

Data set were purified from multivariate and univariate outliers thus 19 of participant data were discarded from the set. Skewness and kurtosis coefficients, and measures of central tendency were calculated, and histogram was analyzed to assess normality of the whole scale and subscales. Since distribution coefficients of scales except academic and artistic ability remained ±1 acceptance range, it was inferred that the data fits to be used by parametric tests (Çokluk, Şekerçioğlu & Büyüköztürk, 2012). When distribution histogram was also examined, similar results were obtained for all subscales. For this reason, it was decided to use parametric tests for cognitive ability and creativity scales. Validity (in content, construct and concurrent dimensions) of the proposed theoretical construct and reliability (in internal consistency, stability and item-test based) of the measurement were inspected. In this sense, EFA and CFA were utilized for construct validity. The Pearson Product Moment Correlation Coefficients between Stanford-Binet Intelligence Test and the cognitive ability subscale, and Torrance Test of Creative Thinking – Figural B (TTCT-B) and the creativity subscale scores were inspected for concurrent validity. Moreover, internal consistency dimension of reliability was analyzed by calculating Cronbach's Alpha Coefficients of each subscales. The Pearson Product Moment Correlation Coefficients between scores of two different measurements were computed for stability size of reliability.

2.5 Limitations

This study is limited in several ways one of which is that utilization of CFA with distinct samples has been proposed by Fabrigar et al. (1999). Data was collected from preschools and kindergartens just around Antalya province due to logistic shortages. Since a limited number of students have been reached, it induced generalization problem. Another issue is about concurrent validity of the instrument. In most of scale development or validation studies (Bakheit, 2015; Lee & Pfeiffer, 2006; Li
et al., 2008; Siu, 2010) school-based scores were usually used, as it is expected, however there were no proper scale to obtain such scores for this study. In Turkish preschool education, development observation forms (DOFs) including five domains namely motor, cognitive, language, social and emotional, and self-care abilities are used to assess and screen children. DOFs are verbally filled instruments so it is not possible to obtain statistical data from DOFs directly. If verbal expressions of teachers could be transformed into artificial discontinuous variable, it would be possible to analyze concurrent validity of GTCEI regarding academic ability. Criterion instrument for artistic ability was also absent.

Absence of national theoretical background for gifted and talented children should be considered as the first and the most important reason why Turkey cannot produce gainful development in education of the gifted.

3. Result

3.1 Contrast validity

Four factored theoretical structure of giftedness has been advocated by the researcher and EFA was utilized to uncover factorial design of the scale items. Before conducting EFA, Kaiser-Meyer-Olkin (KMO) and the Bartlett’s tests were used to determine whether sample size was adequate and appropriate for factor analysis. The KMO measure of sampling adequacy value was .771 which is deemed to be 'middling' or 'sufficiently good' (Çokluk et al., 2012). Furthermore, it was observed that obtained chi-square value was significant ($\chi^2(820) = 2889.595; p< .01$) when results of Bartlett’s Test of Sphericity are taken into consideration.

Principal components analysis to extract factors and maximum variability (varimax) method was selected for rotation. At the end of the analysis, fourteen components having eigenvalues equal to or greater than 1 (EV >1) were observed. Contribution of EV >1 component to total variance is 60.583 %. A great deal of diversity exists in factor extraction methods, however EV >1 rule (Kaiser, 1960) and the Scree Test (Cattell, 1976) were used simultaneously in the study. It is stated in several studies (Costello & Osborne, 2005; Henson & Roberts, 2006) that reporting EV >1 rule solely is not suitable. In this context, it was observed that four factors contribute to the variance significantly, after the fourth factor contribution becomes small and approximately the same (see Figure 1).

![Scree Plot Graph](image)

Figure 1. Scree Plot Graph

Within this frame, it was decided to perform analysis again for four factors, which is found as significant because of the compatibility between presumed and observed theoretical structure. Then, it...
was observed that contributions made to total variance by the first factor is 9.29%, the second factor is 9.11%, the third factor is 8.21%, and the fourth factor is 6.09%, in total 32.71.

Factor loading acceptance level was determined as .32, removal of items with factor loadings less than .32 and items with a cross loading that loads at acceptance level or higher on two or more factors are suggested by Tabachnick and Fidell (2001). EFA was repeated by one-by-one item exclusion, consequently, five items due to insufficient factor loading and three items due to cross loading were discarded from the scale.

It was observed that theoretically presumed design is supported statistically as a result of the analysis (Celik & Yavuz, 2018). Item factor loadings change between .47 and .69 in cognitive ability, .41 and .76 in academic ability, .32 and .64 in creativity, and .36 and .57 in artistic talent subscale. Final contributions made to total variance by the first factor is 9.74%, the second factor is 9.43%, the third factor is 9.34%, and the fourth factor is 6.93%, in total 35.46. According to Çokluk et al. (2012), besides factor loading level criterion, items should meet at least .20 common variance value not to be removed. Even though five items do not satisfy this criterion, these items were remained in the scale because of their originality of measured traits.

Continuing with CFA after EFA is the mostly used way to validate the instrument in the development process where there is a vagueness about which index should be used (Sen et al., 2014). Therefore, all statistical operations were performed by guideline of Çokluk et al. (2012) like in EFA. They recommended to validate the scale by using the ratio of Chi-square per degree of freedom value ($\chi^2/df$), RMSEA, GFI, AGFI, StRMR, NNFI, and CF fit indexes. It is suggested that $p$ value should be checked before controlling index values; however, $p$ value is tolerated in most confirmatory studies since significance of it is normally expected because of big sample size. Chi-square ($\chi^2$) is another fit index assessed not solely but rating degree of freedom $(df)$ in CFA. According to Tabachnick and Fidell (2001) chi-square per degree of freedom less than value of 2 is an indication of perfect fit ($\chi^2/df = 1.85$ where $\chi^2 = 1429.12$ and $df = 772$). Also, other fit indexes (RMSEA = .58, SRMR = .81, NNFI = .86, GFI = .87) yield structural validation of the instrument at higher rates (Brown, 2006; Jöreskog and Sörbom, 1993; Sümer, 2000). Note that several high error variances were omitted because high $T$ values of items were observed.

**3.2 Concurrent validity**

Level of relationship between scores obtained by well-established instrument and the another can be used an indicator of this type of validity. It was planned firstly to rate same group of children with Turkish version of GRS-P and GTCEI however accessibility of Turkish version of GRS-P precluded it. For this reason, concurrent validity of cognitive ability and creativity subtests can be analyzed. Stanford-Binet intelligence test (SB) and Torrance Test of Creative Thinking Figural B (TTCT-B) were used as criterion instruments for cognitive ability and creativity subtests orderly. As it is expected, students who scored higher in criterion tests obtained higher scores in GTCEI. Pearson moments correlation coefficient between total scores of SB and cognitive ability was .84, and TTCT-B and creativity is .83, which shows strong evidence for concurrent validity.

**3.2.1 Reliability (Consistency)**

The Cronbach’s alpha coefficient was calculated for each subscale to provide an indication of the internal consistency reliability. The Cronbach’s alpha for ‘cognitive ability’ is .81, for ‘academic ability’ it is .79, for ‘creativity’ it is .75, for ‘artistic talent’ it is .65. For all four scales, the Cronbach’s alpha is .80, greater than .70, suggesting that these scales had satisfactory reliability in assessing giftedness in preschoolers. Additionally, item-test statistics showed that corrected item-total correlations ranged from .25 to .66 for whole scale. It was also observed that deletion of any item diminishes Cronbach’s Alpha value of the scales.

**3.2.2 Reliability (Stability)**

A series of Pearson moments correlations were calculated to examine test-retest reliability of the data obtained by cognitive ability, academic ability and creativity subscales. The participants' first
measurement scores were related to second ones in all subscales orderly ($r_{xy} = .95; .93; .89$, $p < .01$). For the same end, Spearman's Rho was calculated for the data obtained by artistic talent subscale, and there was a significant positive relationship between the two measurements ($r_s = .86$, $p < .01$). These values clearly indicate strong test-retest reliability of the data obtained by GTCEI.

4. Discussion

This study was aimed at developing a new scale by which possible gifted and talented students can be nominated for the following steps of identification at the maximum preciseness and investigating psychometric properties of it. A series of statistical computations and operations were used to elucidate to what extent obtained data by GTCEI valid and reliable are. The findings showed that GTCEI is a useful instrument which has robust psychometric features consisting of four-factor structure, high construct and criterion-based validity, internal consistency and stability. One issue should be considered that internal consistency coefficient belonging artistic talent is relatively low. This statistical deficit probably results from that prospective teachers in universities do not formally engage in artistic endeavors unless they have specific interest towards art. Unlike other domains in the scale, artistic insight can be obtained by learning heuristics of art if inborn supreme talent is not the issue (Engudar, Sarioğlan & Dolu, 2020). In other words, how to make art, and therefore evaluating it, can be made possible only by engaging in artistic processes or making art. What is more, invitation of another teacher who is more familiar with students’ artistic skills is suggested by Pfeiffer & Jarosewich (2003) while students are being rated. Domain expertise is needed, evaluation of artistic skills should be performed by more ‘educated eyes.’

There has been a propensity for inclusion of psychological and environmental variables into intellectual ones in contemporary giftedness models in which the talent undergoes a transition from potential to eminence (Stoeger, 2015). Some researchers (e.g., Feldman, 2000; Gagne, 2005; Tannenbaum, 1986) integrate them in their models or works because they claim that these components play a binding role in talent development. However, it is hard to make an inference in detail and observe latent variables underlying giftedness by using a single rating-based nomination scale. For this reason, environmental elements, like motivation, is not regarded as a type of giftedness itself, therefore not included in GTCEI.

The prime challenge in the study was making a choice between adopting previously presented structure of giftedness, then adapting its instrument into Turkish culture and developing a new self-presumed one. The latter was found more suitable and the rationales for why a new scale was developed show exact consistencies with study statements of Humbleton and Patsula (1999). Firstly, they stated that adaptation of existing test should not always be preferred instead of developing a new test for a second language group, when making a comparison among different cultures is not of interest. The researcher gave priority to recognizing developmental indicators of preschoolers acknowledged by Turkish MoNE and strived to just be inspired by other measurement tools. Secondly, choice of developing new scale allowed the researcher to make desired modification and improvements thanks to feedbacks acquired by experts and teachers. Removal of two items after the pilot study can be shown as an example of this stance. Once the researchers or contributors are not familiar with source and target languages and the cultures, translation problems rise (Ozbey, 2020). Because it is hard to reach a vast number of field experts who spend their times in both cultures in Turkey, developing a new scale was the reasonable choice.

The scale is different from its world-wide counterparts in several ways one of which is that it comprises examples of items. When it comes to understanding theoretical indicators in a meaningful way, unfamiliar words or concepts prevents raters whose subject-matter and pedagogic adequacy are controversial from assessing the subjects fairly (Cosar, Çetinkaya and Çetinkaya, 2015; Sedghi & Najafi, 2018). Giving examples of items decreases complexity, makes expected behaviors more concrete and eases use of instrument. Concordantly, teachers who also attended the pilot study welcomed examples of items and reported an instrument consisting of items with examples serves well, and even better than the initial form of it.
What is more, the scale is culturally compatible. Giftedness is a notion which can have immense varieties according to where it conceptualized and constructed. It is clearly stated opinion by many researchers (e.g., Hertzog, 2008; Pfeiffer & Blei, 2008) that giftedness is culturally dependent and affected by its belonged culture, and even the types of giftedness offered by school systems may show substantial variations in different school districts. This discrepancy shows itself among rating scales, for instance, Gifted Rating Scales – Preschool / Kindergarten Form includes items respecting reading and writing skills which are not expected from preschoolers in Turkey. Teachers reported during off-record conversations that although some signs of reading or writing have been observed by them in leisure time activities, it would not be a realistic rating if they score reading and writing skills of their students.

5. Conclusion

Size and representativeness seem two major concerns about the sample used in the study. Number of participants should be increased, and more attention should be paid to demographic characteristics of the sample to be able to generalize the results to all preschoolers in Turkey. Additionally, whether participants differentiate significantly according to age bands and gender can be investigated in further research. On the other hand, it can be firmly claimed that cultural differences affect how the giftedness is depicted. There may or must be no agreement among definitions of giftedness in different locations. Nationally and specific to general, definition of giftedness should be conceptualized throughout an evolution process from applicable understanding reflecting cultural norms to nationally accepted construction.

References:


