Klein, B., & Fodor, S. (2019). *TalentTiles:* A new descriptive talent identification instrument based on teachers' ratings. In R. F. Subotnik, S. G. Assouline, P. Olszewski-Kubilius, H. Stoeger, & A. Ziegler (Eds.), *The Future of Research in Talent Development: Promising Trends, Evidence, and Implications of Innovative Scholarship for Policy and Practice. New Directions for Child and Adolescent Development*, *168*, 11–25.

2

TalentTiles: A New Descriptive Talent Identification Instrument Based on Teachers' Ratings

Balázs Klein, *Szilvia Fodor*

Abstract

Talent *identification based on multiple criteria is a common way of considering students for gifted education services. As part of a nationwide talent identification program led by UNK (New Generation Centre, Hungary of Talents Program, 2016–2020), we developed an online rating tool with which teachers could indicate the talent areas of their 5th grade students. Through several iterations—involving the translation, application, and analysis of Renzulli's Scales for Rating the Behavioral Characteristics of Superior Students questionnaire and a custom-made normative questionnaire—we developed a matrix-style competency rating instrument called TalentTiles. The primary aim of the instrument is to collect more realistic and reliable teacher-ratings on students. The instrument takes a very short time—less than 2 minutes per student—to complete, focuses on the behavioral characteristics, measures school-related areas, addresses general competencies (e.g., ability, motivation) in a domain-specific way and has low average inter-item correlations. After the first pilot test TalentTiles seems to be an innovative and user friendly way of collecting teachers' opinions about school-related gifted behavior.* © *2019 The Authors.* New Directions for Child and Adolescent Development *published by Wiley Periodicals, Inc.*

New Directions for Child and Adolescent Development, no. 168, November 2019 © 2019 The Authors. *New Directions for Child and Adolescent Development* published by Wiley Periodicals, Inc.
Published online in Wiley Online Library (wileyonlinelibrary.com). • DOI: 10.1002/cad.20317

Teachers' Evaluation as a Tool for Talent Identification

The Advantages of Rating Scales. Accurate identification is an important step when planning services for gifted students, although a survey of experts highlights the challenges we face in the field of identification, especially methods and processes (Pfeiffer, 2003). Over the past few decades a great deal of debate emerged over the appropriateness of using teacher rating scales in the identification of gifted students. Teacher ratings and nominations have often been criticized for their lack of validity (Pegnato & Birch, 1959), although some other studies, such as the argumentation of Gagné (1994) shows that teacher ratings are not worse than most other sources of information.

Employing multiple criteria—including teachers' ratings and test score information—is the most common way of considering students for gifted education services. Different assessment tools provide different information about the characteristics we associate with giftedness. The logic of using multiple criteria is to rely on more sources of information to improve the selection process. The benefit of rating instruments is obvious: they can quickly and with relatively little effort provide qualitative information on the behavior of students that would be difficult to obtain otherwise. In addition, this method is relatively cheap, goal-oriented, and focused, and students' behavior can be observed in their everyday and natural environment.

The success of teachers as raters stands out when teachers rate explicit behaviors as opposed to providing opinions. Most validity studies suggest that teachers can be successful in identifying students, at least when correlated with outcomes such as achievement and ability tests (Peters & Gentry, 2012). On the one hand, correlations between teacher-rating scales and achievement tests are usually moderate, but on the other hand, if these correlations were too high, then the ratings would simply duplicate the information of the achievement tests without contributing anything new.

The Limitations of Rating Scales. While there are unquestionable benefits of rating scales, teachers' assessments are inevitably subjective. This subjectivity comes with some limitations that must also be considered:

- Filling out a questionnaire requires the teachers' motivation and familiarity with the student,
- There is great variation between teachers' scoring style,
- The evaluation focuses heavily on school-related aspects of giftedness, leaving other important domains out of consideration.

Teacher's evaluations or recommendations are often considered to be the cheapest, and for this reason used as the first step of identification before more expensive types of assessment, like objective testing. In our opinion, both the reasoning and the effect of this practice are flawed. Teachers'

recommendations as a first filter should be avoided unless the aim is specifically to identify well-adapting high achievers for the following reasons:

1. Today objective online testing can be very cost-effective, too, especially when compared to the implicit cost of teachers' time. Combining these two methods of assessment, a complex picture of students can be gained in an economic way.
2. Depriving students of the chance to show their strengths on objective tests based on teachers' subjective opinion prevents the identification of underachievers. Identification systems that require teacher nomination before testing results in a large proportion (around 60%) of gifted students being missed (McBee, Peters, & Miller, 2016).

Below is a short sample of the limitations of rating scales broken down according to the ORCE (Observe, Record, Categorize, Evaluate) approach designed to help counter observer biases in assessment contexts in organizations (Woods & West, 2010).

- *Observation*. Often the assessors can only observe the focus person in very specific settings (e.g., in class) which can limit the range of displayed behavior by the focus person.
- *Recording*. Out of the displayed behaviors of the focus person, the assessor will only note a small portion at the time that it happens and will recall an even smaller portion when the questionnaire is filled. Attention and memory filters and distorts our recollections—and since both of these mechanisms are known to be biased by several factors—they can yield a very subjective outcome.
- *Categorization*. Without thorough and standardized trainings the categorization of behaviors can be problematic. The scales of the questionnaires must indeed be very clearly defined to help observers categorize behavior as evidence of "creativity," "leadership," or "motivation."
- *Evaluation*. Whether a behavior is out of the ordinary (e.g., significantly more frequent or more skilled) than that of the others can also be a matter of subjective—and result in a possibly biased—judgement.

The Development of TalentTiles

Having looked at the advantages and disadvantages of teacher rating scales in general we decided to create an instrument that takes all these concerns into consideration and offers a professionally designed and evidence-based solution for collecting and evaluating teachers' observations. The process of creating the new instrument had three steps:

a) *Translation and Trial of the Renzulli Scale*. First the well-known and widely used measure, the Scales for Ratings of the Behavioral

Characteristics of Superior Students (SRBCSS: Renzulli et al., 2010) was translated and trialed with 500 students.

b) *Creation and Trial of Our Own Teacher Rating Questionnaire*. Because of the unsatisfactory statistical results of the SRBCSS, we created a completely new questionnaire to find out the relevant domains of observation. This questionnaire was filled out by 32 teachers on 268 students.

c) *Creation and Trial of TalentTiles*. Based on the above results an innovative, matrix-style instrument was designed and called TalentTiles. The instrument has undergone a recent pilot study ($n = 114$) and provides promising results about collecting teachers' opinion about behavioral signs of giftedness.

Translation and Trial of the Renzulli Scale. Among the more than thirty assessment scales used on the international scene, Scales for Ratings of the Behavioral Characteristics of Superior Students (Renzulli et al., 2010) is one of the most well-known. Originally developed in the 1970s, it has been continuously translated into several languages making it useful for comparative international studies of school talent. The first three scales of the questionnaire (learning, motivation, creativity) were developed to guide identification based on the three-ring conception of giftedness (Renzulli, 2005). As the conception of giftedness began to expand, additional scales were added to the questionnaire to examine high potential, and now it has fourteen scales measuring Learning, Creativity, Motivation, Leadership, Artistic, Musical, Dramatics, Communication-precision, Communication-expressiveness, Planning, Mathematics, Reading, Technology, and Sciences talent characteristics.

In the first step of our study, the questionnaire was filled on 500 5th grade, primary school students and data were analyzed together with the students' performance on two ability tests (fluid intelligence, vocabulary) and their average school marks (AvgMark) of the previous year.

The two ability tests are both online adaptive tests using item-response scoring.

- SAM (Scrambled Adaptive Matrices) measures fluid intelligence—or problem solving ability—in a problem domain similar to the Raven's Progressive Matrices (Klein, Raven, & Fodor, 2018),
- NoVo is a vocabulary scale measuring crystallized intelligence, or generic knowledge about the world (Raven, 2000).

Some of the correlation results support the validity of the SRBCSS scales, while others are more difficult to interpret. Table 2.1 shows the significance level of the different SRBCSS scales in predicting SAM, NoVo, and AvgMark results:

Table 2.1. Significance Level of Linear Regression Coefficients While Predicting Different Performance Measures From the Renzulli Scales ($n = 500$)

	SAM	*NoVo*	*AvgMark*
Creativity	*	***	
Drama	*	**	
Learning	*		*
Mathematics	**		***
Motivation	***		***
Technology	*		***
Reading			***
Science			***
Leading			**
r	**0.5**	**0.6**	**0.7**

Significance level: * – 1%, ** – 0.1%, *** – 0.

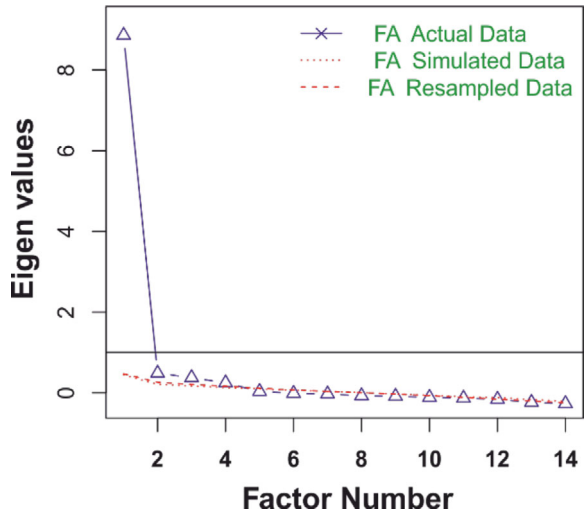
- Students good in problem solving are seen to be motivated and skillful in mathematics,
- Students with a large vocabulary are seen as creative and good in drama,
- Students with good grades are seen to be motivated and good in leading others, as well as good in mathematics, technology, reading, and science.

Out of the three performance variables (problem solving, vocabulary, grades) average grade could be best predicted ($r = 0.7$) from the teachers description, when using the Renzulli scales. This suggests that ratings on the Renzulli scales are closely related to general performance at school.

Similar to previous findings (Gridley & Treloar, 1984), our analysis showed a very high level of intercorrelations ($r = 0.68$ on the average) between the fourteen scales of the questionnaire. To determine the number of factors, a parallel analysis technique of the "psych" package in R was used. Sharp breaks in the plot suggest the appropriate number of components or factors to extract (Revelle, 2017). As shown in Figure 2.1, factor analysis suggested only one single factor. When interpreting the results, we agree with previous authors (Kaufman, Plucker, & Baer, 2008) that the practice of using SRBCSS scales for measuring separate sets of characteristics is questionable.

Together with their ratings of students, teachers' ($n = 29$) subjective opinion about the rating process itself was also collected. Their critique addressed two main concerns: the 126-item questionnaire took too much time to complete, and it required them to report on areas of the children's life about which they had no information. This feedback helped us to formulate a few hypotheses to explain the possible reasons of the high intercorrelations and the unidimensional factor-structure:

Figure 2.1. Unidimensionality of the Renzulli Scales ($n = 500$).



- Teachers find it too demanding and time-consuming to focus on each question of the questionnaire. Giving uniform—or very similar-answers can save a lot of time and energy for them.
- Teachers formulate a general impression of the student's abilities, and they are using this impression to fill in the questionnaire in a wide range of areas. Teachers of upper graders (5th–8th grades) meet students only when teaching their particular subject, a few hours a week, so they generalize from what they have experienced in their own lessons or because of the generic halo effect (Thorndike, 1920).

Creation and Trial of Our Own Teacher Rating Questionnaire. As it turned out, most of the teachers did not have the necessary information about students to characterize them in all the talent domains, so it was necessary to define, what domains they are able to observe and identify at school and what areas we should include in our new rating instrument. To identify the relevant areas and the observable behaviors, a totally new normative questionnaire was created measuring fifteen talent areas (Cognitive abilities, Motivation, Creativity, Leadership, Mathematics, Sciences, Hungarian language, Foreign language, Technology-informatics, Spatial-visual, Artistic-visual, Sport-kinesthetic, Dance, Music, and Drama-performance skills).

First, items and statements describing above-average behavior in the given fifteen talent domains were generated. These items were judged by subject matter experts by classifying the items into the best fitting category and identify hard-to-classify items. During this process, it transpired that

some generic areas (such as Cognitive abilities, Motivation, Creativity, Leadership) are hard to interpret separate from a domain. A child—for example—is not motivated in general, but motivated to do something specific (e.g., motivated to go to a math competition, or motivated to listen to music in his/her free time). Similarly, a statement like "Displays perseverance and concentration while dealing with music" could be categorized either to be a statement about Motivation or about Music. Therefore, we decided to carry out a major change in the structure of the questionnaire: the competency-related, general statements of Ability, Motivation, Creativity, and Leadership were incorporated into the specific, subject-related scales, so in each subject-related scales (Sciences, Hungarian language, Foreign language, Technology-informatics, Spatial-visual, Artistic-visual, Sport-kinesthetic, Dance, Music, and Drama-performance) there were statements about domain-specific Ability, Motivation, Creativity, and Social behavior. Finally, we ended up with an instrument measuring 11 areas with 158 statements.

Next, the questionnaire was filled out by thirty-two teachers on 268 students. After performing a factor analysis, seven scales remained. Drama, Spatial-visual, Dance, and Sciences did not form separate factors in our data. A possible reason could be that Drama, Dance, and Spatial-visual skills are not separate subjects in the Hungarian educational system, so teachers cannot really observe them in their own lessons. Sciences are—of course—included in the system, although it starts as a separate subject only at the 5th grade, so teachers do not really have the opportunity to observe science-related behaviors. Therefore—as the target group of our questionnaire is the 5th graders—these scales were not included in the final version of the instrument as separate areas. The remaining seven areas were the following:

- Mathematics (Mat)
- Motion (Mot)
- Music (Mus)
- Information technology (It)
- Visuals (Vis)
- Hungarian language and literature (Lit)
- Foreign language (For)

These seven scales proved to be well defined, showed sufficient reliability and acceptable intercorrelations as described in Table 2.2.

Creation and Trial of TalentTiles. Based on the experiences about the Renzulli Scales and other normative teacher rating questionnaires, six principles guided the development of the *TalentTiles* instrument. We decided that it has to be:

- Scientifically sound (based on relevant studies and verified against external criteria),

Table 2.2. Intercorrelation Between the Scales of Our Custom Rating Scale ($n = 268$)

	Mat	*Mot*	*Mus*	*It*	*Vis*	*Lit*
Mot	.30					
Mus	.33	.41				
It	.62	.56	.26			
Vis	.40	.41	.54	.46		
Lit	.68	.45	.42	.68	.56	
For	.70	.37	.33	.61	.45	.70

- Behavior-focused (addresses phenomena that can be directly observed by the teachers),
- Domain-specific (allows teachers to focus on areas they are familiar with, while not forcing them to report on areas they are not),
- Positive, strength-oriented (searches for existing talent areas independently of possible weaknesses in other areas),
- User-friendly (short, requires minimal effort to complete),
- Differential (look for specific behaviors instead of a general picture).

From the list above, the term "differential" merits some further explanation. Our goal was to get a much lower average intercorrelation between the items than it is usual on traditional normative questionnaires, so we looked at ipsative instruments. An instrument is ipsative when raw scale scores sum to a constant for any individual (Baron, 1996). Ipsative instruments are used for two main reasons:

- To better control of response patterns (like central tendency, or social desirability) and
- To reflect the position that life is about choices.

Ipsative scales reduce, while normative scales increase the intercorrelation in the data (Saville & Willson, 1991). Ipsative instruments show which traits are strongest and weakest in the individual, but not the absolute value of scale positions. Since previously—on both the Renzulli scales and our custom questionnaire—we found high correlations between the scales, when designing the new instrument some level of ipsativity seemed to be advisable, especially considering that our purpose with the questionnaire was purely a description of the children and not—in any way—ranking them.

Having identified the seven relevant behavioral areas, appropriate structure and item-arrangement was needed to meet our previously stated goals. Our previously created normative questionnaire already had an implicit matrix-style arrangement: each area had items on four different competencies. It therefore made sense to transform that to an explicitly

Figure 2.2. The *TalentTiles* instrument.

	Skill	Interest	Perseverance	Creativity	Support	Leadership
Mathematics	☐	☐	☐	☐	☐	☐
Motion	☐	☐	☐	☐	☐	☐
Music	☐	☐	☐	☐	☐	☐
IT	☐	☐	☐	☐	☐	☐
Visuals	☐	☐	☐	☐	☐	☐
Literature	☐	☐	☐	☐	☐	☐
Language	☐	☐	☐	☐	☐	☐

matrix instrument similar in concept to that of the Edinburgh Questionnaire (Raven, 1997). We arranged the seven topic areas as rows and the four components—Ability, Motivation, Creativity, and Social behavior—as columns. The cells of the online instrument contain tick boxes by which teachers can indicate exceptional behavior that they have observed. While creating this matrix-style arrangement, it transpired that in some cases it is difficult to give consistent description to Motivation and Social behavior; therefore, we split both of these scales into two: Motivation into Interest and Perseverance, Social behavior into Leading and Social Supporting.

To signify the matrix arrangement of the instrument we named it **TalentTiles**.

TalentTiles consists of a matrix with seven relevant content domains (Mathematics, Motion, Music, Information technology, Visuals, Hungarian language and literature, Foreign language) as rows, and six different competency components (Ability, Interest, Perseverance, Creativity, Support, Leadership) as columns, shown in Figure 2.2. The task of the teachers is to mark two to eight cells describing an above-average behavior, which can be observed in the given student. The instrument fits on a single screen and requires only two to eight clicks to complete with minimal amount of reading, thus making it very quick to complete. It has a limited number of relevant domains, looks only on positive, observable behaviors and forces raters to make choices, while allowing them sufficient autonomy in decision-making.

Figure 2.3. An example for a *TalentTiles* description.

Based on your description, the child you rated …

Finds it easy to comprehend complicated mathematical content. Is interested in working on mathematical problems. Works persistently on mathematical problems. Finds interesting, novel and unusual solutions to mathematical problems. Helps her peers by explaining complicated mathematical topics. Takes leadership in groups working on mathematical problems.

Each tile is associated with a single sentence description, which—if selected—is automatically added to a narrative description. With further selections these sentences are collected into a short description. This description is constantly displayed and updated real-time. We had a double purpose with this method: we intended to give raters immediate feedback on their choices and also provide an easy-to-read narrative result to other stakeholders.

A narrative description where all Mathematics related components were selected would look like Figure 2.3.

Empirical Results on TalentTiles

With the aim of examining *TalentTiles*, a pilot study was carried out in a Hungarian public elementary school with 5 teachers, who submitted 114 ratings on 97 children.

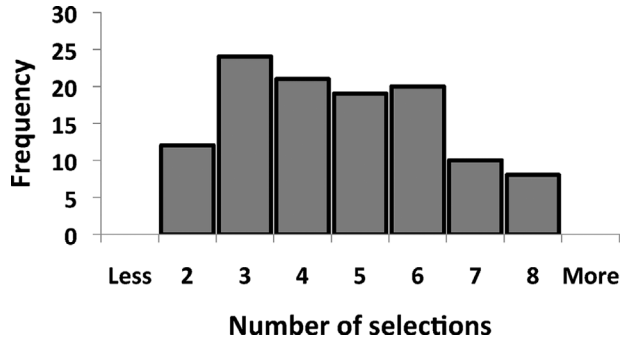
Response Times. The average time for a teacher ($n = 5$) to rate a student—when they first were introduced the instrument—was less than 8 minutes, subsequent ratings, however, took only about 1.5 minutes (1.4 minutes, $n = 108$, one outlier removed) on the average. These results show that first teachers have to familiarize themselves with the instrument—understand its structure and layout—but subsequent ratings can be conducted very quickly.

Number and Distribution of Selected Behaviors. When examining the distribution of selections in each cell—as shown in Table 2.3—it can be seen that certain areas are much more frequently selected than others. This could mean that the rated students were indeed more talented in those areas, or—more probably—that teachers noticed and appreciated those above-average behaviors more often. The competencies of Interest, Skill and Perseverance are more often selected than Creativity, Support, and especially Leadership. This is probably due to the unfortunate educational environment where—with the lack of group work and self-motivated

Table 2.3. Total Number of Selections ($n = 114$) (row and column sums contain average number or selection per rated students)

	Skill	*Interest*	*Perseverance*	*Creativity*	*Support*	*Leadership*	*Areas*
Mathematics	18	14	8	2	7	5	**0.47**
Motion	31	47	22	7	3	0	**0.96**
Music	10	32	2	1	0	0	**0.39**
IT	20	25	12	3	1	0	**0.54**
Visuals	28	43	11	13	10	3	**0.95**
Literature	34	39	21	11	14	6	**1.10**
Language	7	14	4	0	1	0	**0.23**
Components	**1.30**	**1.88**	**0.70**	**0.32**	**0.32**	**0.12**	**4.64**

Figure 2.4. Distribution of total selections ($n = 114$).



activities—these latter competencies have rare opportunities for display and are therefore difficult to observe. Literature, Motion, and Visual categories were the most frequently rated, suggesting that these were the subjects the observers taught in their classes.

Requiring raters to make genuine choices and preventing them from giving uniformly positive or negative responses, raters were instructed to select minimum two and maximum eight cells. Looking at the distribution of the number of selections they made, we could detect no obvious floor or ceiling effect (Figure 2.4). Since it probably would not change the number of selections significantly there is nothing against allowing teachers to make more than eight selections, on the other hand it seems desirable to keep the requirement of at least two selections for each child.

Intercorrelations of Items. Though *TalentTiles* is technically not an ipsative instrument (the total number of ticked checkboxes does not have to be, and is not the same for everyone), its correlation matrix shows remarkable similarities to that of ipsative instruments. Figure 2.5 shows the intercorrelations of the items of SRBCSS and the *TalentTiles*. Although the average intercorrelation of items is quite high (0.59) for the SRBCSS, it

Figure 2.5. Distribution of item-correlations in SRBCSS ($n = 500$) and TalentTiles ($n = 114$).

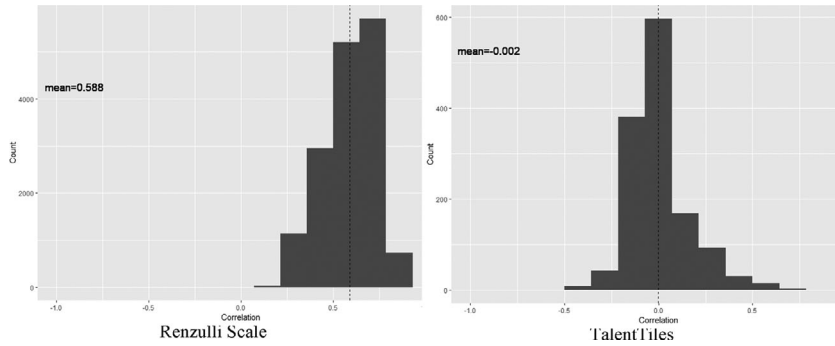


Table 2.4. Cronbach Alphas of the Areas and Competencies of TalentTiles ($n = 114$)

Areas	*Alpha*	*Competencies*	*Alpha*
Mathematics	0.65	Skill	0.34
Motion	0.65	Interest	0.08
Music	0.08	Perseverance	−0.32
IT	0.69	Creativity	0.00
Visuals	0.58	Support	−0.15
Literature	0.82	Leadership	0.22
Language	0.13		

is practically 0.00 for the *TalentTiles*. This latter result is typical to ipsative questionnaires.

Our assumption is that in our case selection limits has only minimal effect on the results and ipsativity occurs mainly because of the test format. Teachers are only willing to invest a finite amount of effort into a rating and each response—ticking a box—consumes some of this effort. This keeps the amount of selections low and creates a feeling that they need to choose between selection options. This reasoning allows us to formulate a further hypothesis: we think that increasing—or even totally releasing—the constraint on the maximum number of allowed selections (currently set to 8) will not increase average intercorrelation with a significant amount. This hypothesis—of course—needs to be tested in further trials.

Domains and Competencies. Although our aim with TalentTile was to create a purely descriptive instrument, we examined how reliable its scales were if we created them as a sum of selections across rows (as talent-domains, like Math, Music, and others) or columns (as competencies, like Skills, Motivation, and others). Table 2.4 shows the results of reliability-tests of these scales.

It can be seen that most of the Domains form reliable scales. The two exceptions are Music and Language. We assume that teachers involved in the nomination process did not teach these subjects, so they did not have the opportunity to observe students in these areas.

Competencies of the instrument do not form reliable scales if we add up selections across domains: there is no, or only low correlation between the competency-items of different domains. This suggests that competencies are not general, but rather domain-specific. For example, the negative alpha of Perseverance suggest that students are not hard-working or lazy in general, but perseverance in one domain can reduce the chance of being perseverant in another domain (e.g., a lot of time spent on Music can reduce the time spent on Mathematics).

Altogether, while—with due care—it could be possible to create standardized scores from domain responses of *TalentTiles*, it is not possible to create aggregate scores of competencies and—certainly in this regard—the instrument should only be used in a descriptive manner and not as a tool to create rank orders. This is in accordance with our stated goals.

Summary

Teachers are in a very good position to identify different talents of their students, but creating instruments to formalize this process has proved to be a difficult task. First, in this project, the Renzulli's Scales for Ratings of the Behavioral Characteristics of Superior Students were translated into Hungarian and piloted with 29 teachers and 500 students at the age of 10–11. Teachers found the questionnaire too long to complete and results showed high intercorrelations ($r = 0.68$ on the average) among the scales. In other words—in our case—the quality of the obtained information could not justify the amount of effort it took to complete the SRBCSS.

Trying to overcome these shortcomings, we have developed a new rating instrument called *TalentTiles* with the aim of gaining a limited amount of very clearly defined information in a very quick but reliable manner. This online instrument consists of a matrix of relevant subject domains as rows (Mathematics, Motion, Music, Information technology, Visuals, Hungarian language and literature, Foreign language) and six different competency components (Ability, Interest, Perseverance, Creativity, Support, Leadership) as columns. Teachers can indicate observed talent areas by selecting the checkboxes in the appropriate cells. In a subsequent pilot study, 114 ratings were collected, and according to empirical testing, *TalentTiles* was found to be very quick to complete and provide differential results similar to that of ipsative instruments. The reliability analysis of teachers' nominations show that scales could be created by aggregating ratings within domains, but competencies are entirely domain-specific and cannot be aggregated into scales. It supports the domain-specific view of giftedness (Subotnik,

Olszewski-Kubilius, & Worrell, 2011) and highlights the importance of creating talent-identification instruments with a domain-specific focus.

It also turned out from the personal feedbacks that the rating of explicit classroom behavior is easier for teachers than providing general opinion, because it gives clearer instructions and guidelines for them and focuses their attention, which is in line with Peters and Gentry's (2012) findings. Compared to the questionnaire-form teacher rating scales, *TalentTiles*, with its matrix-style arrangement and ipsative structure, facilitates teachers' decision-making, therefore concentrates on differences between students and elicits focusing on individual strengths. So our original goal of creating a scientifically sound, behavior-focused, domain-specific, strength-oriented, user-friendly, and differential instrument was achieved.

At the same time, there are some limitations of our study that derive mainly from the small sample size and the lack of a thorough validation process. It is obviously necessary in the future to continue the examination of *TalentTiles* with a much higher number of students and teachers, and to validate our results by comparing our results to other measures like ability tests and student achievement. It is also worth collecting data about students' behavior not only from teachers, but also from parents and students themselves, and to compare different teachers' opinion about a given student, which may provide a much richer understanding of gifted classroom behavior.

References

Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, *69*(1), 49–56.

Gagné, F. (1994). Are teachers really poor talent detectors? Comments on Pegnato and Birch's (1959) study of the effectiveness and efficiency of various identification techniques. *Gifted Child Quarterly*, *38*(3), 124–126. https://doi.org/10.1177/001698629403800305

Gridley, B. E., & Treloar, J. H. (1984). The validity of the scales for rating the behavioral characteristics of superior students for the identification of gifted students. *Journal of Psychoeducational Assessment*, *2*(1), 65–71. https://doi.org/10.1177/073428298400200108

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. Hoboken, NJ: Wiley.

Klein, B., Raven, J., & Fodor, S. (2018). Scrambled Adaptive Matrices (SAM)–A new test of eductive ability. *Psychological Test and Assessment Modeling*, *60*(4), 441–482.

McBee, M. T., Peters, S. J., & Miller, E. M. (2016). The impact of the nomination stage on gifted program identification. *Gifted Child Quarterly*, *60*(4), 258–278. https://doi.org/10.1177/0016986216656256

Pegnato, C. W., & Birch, J. W. (1959). Locating gifted children in junior high schools. *Exceptional Children*, *25*(7), 300–304. https://doi.org/10.1177/001440295902500702

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

Pfeiffer, S. I. (2003). Challenges and opportunities for students who are gifted: What the experts say. *Gifted Child Quarterly*, *47*(2), 161–169. https://doi.org/10.1177/001698620304700207

Raven, J. (1997). *Competence in modern society: Its identification, development and release*. Unionville, NY: Royal Fireworks Publishing.
Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*(1), 1–48.
Renzulli, J. S. (2005). The three-ring conception of giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 246–279). Cambridge, UK: Cambridge University Press.
Renzulli, J. S., Smith, L. H., White, A. J., Callahan, C. M., Hartman, R. K., Westberg, K. W., … Systma Reed, R. E. (2010). *Scales for rating the behavioral characteristics of superior student: Technical and administration manual* (3rd ed.). Mansfield Center, CT: Creative Learning Press, Inc.
Revelle, W. (2018). *psych: Procedures for personality and psychological research*, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psychVersion=1.8.12. https://cran.r-project.org/web/packages/psych/citation.html
Saville, P., & Willson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, *64*(3), 219–238.
Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056
Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, *4*(1), 25–29.
Woods, S. A., & West, M. A. (2010). *The psychology of work and organizations*. Boston, MA: South-Western Cengage Learning.

Dr. Balázs Klein is managing director of a Budapest based assessment company (Testar Ltd.) and a researcher at Labor and Health Sciences, University of Pécs, Hungary. His research interest is computer based talent identification and assessment in educational and work context. His online, IRT based intelligence test (SAM) is used in several large scale projects worldwide.

Dr. Szilvia Fodor is an assistant professor at the Department of Educational Psychology, University of Debrecen, Hungary, with a research interest in talent identification, creativity and school climate. In addition to academic work, she contributes to several talent support projects, like the Hungarian Templeton Project (2015–17) and the Hungary of Talents (2016–20), is employed by the European Talent Centre — Budapest, and holds a course at the European Leadership Program at the Jesuit Eusopean Social Centre in Brussels.