Measuring metaphors: Concreteness and similarity in metaphor comprehension and gifted identification

Mei Tan, Baptiste Barbot, Catalina Mourgues & Elena L. Grigorenko

In this article, we present a subtest of the Aurora Battery, Metaphors*, a new measure for the identification of gifted students.* Metaphors *is an assessment of analytical thinking in the verbal domain with an open-ended response format, an alternative to the conventional achievement measures or analytical assessments in multiple-choice format often used for giftedness identification. The comprehension of metaphors requires a combination of comparative, categorical, and evaluative thinking, and an ability to map meanings – literal and figurative – from one word to another. In this article, we first show that performance on* Metaphors *can contribute to the identification of giftedness, and we identify the items that seem to contribute to this discrimination most effectively (Study 1). Next, the comparative terms of each subtest item were characterised according to two key aspects – the concreteness and similarity of their comparative terms (Study 2). Results showed that gifted children may be better identified when these aspects are considered. New perspectives on metaphor processing in gifted students and premises for the identification of giftedness are then discussed.*

Keywords: *Giftedness; assessment; analytical thinking; metaphors.*

METAPHORS are complex linguistic and cognitive constructions; hence the quality of students' comprehension of such constructions may contribute to the differentiation of gifted from non-gifted children. Metaphors present identity statements that are not meant to be understood literally but figuratively, for example: 'Life is a journey,' (i.e. 'X is a Y.'*). Metaphor comprehension requires a combination of comparative, categorical, and evaluative thinking, as well as an ability to map meanings – literal and figurative – from one word to another, establishing correspondences between words that may differ widely (Gentner & Bowdle, 2008; Gibbs, 2011; Glucksberg, 2008; Lakoff & Johnson, 1980; Sternberg & Nigro, 1983). Understanding metaphors, then, taps into a form of analytical thinking in the verbal domain, a cognitive component frequently considered when evaluating levels of giftedness (e.g. using standardised IQ tests). Historically considered to be under the purview of literary study as examples of poetic language, metaphors are now understood to be natural components of everyday language, unconsciously used and often quickly comprehended. They represent essential aspects of cognition, knowledge-structuring and communication (e.g. Evans, 2010; Gibbs, 2011; Gibbs & Tendahl, 2011; Kovecses, 2010; Lakoff & Johnson, 1980), yet the quality and complexity of their production and comprehension may differ across individuals. Thus, metaphors have become an active subject of research in various fields such as linguistics, psychology, political science and philosophy (see Gibbs, 2008; Ortony, 1979b, 1993, multidisciplinary compendia of metaphor research).

* In the literature on metaphors, the X/Y terms are variously called the 'tenor/vehicle' or 'target/source or base' (e.g. Life/Journey); in this article, we use the terms 'target/source' as being more current. The ways in which the two terms are related is called the 'ground' of the metaphor.

A variety of theories of metaphor processing proposed over the years emphasise aspects of metaphor structure and comprehension that are relevant to the analytic mechanisms elicited by *Metaphors*' items. First, the initial phase of metaphor comprehension involves defining the relationship between the target and source terms. This occurs through an analogy-like comparison of X and Y term attributes (e.g. Evans, 2010; Gibbs, 2011; Gibbs & Tendahl, 2011; Kovecses, 2010; Lakoff & Johnson, 1980), or the discernment of the superordinate category that contains both terms (Gentner & Bowdle, 2008; Gentner & Markman, 1997; Gentner & Namy, 1999; Sternberg & Nigro, 1983; Tourangeau & Rips, 1991; Tourangeau & Sternberg, 1982). Second, metaphor comprehension is strongly influenced by an inherent directionality in metaphor structure, such that it is generally the source term (Y) that tends to dominate or define the characterisation of the comparison, especially when the source and target are less similar (Glucksberg, 2008; Glucksberg & Keysar, 1990). Third, metaphors must make sense, and, therefore, the process of comprehension includes an evaluation for some degree of logical sense or meaning (comprehensibility), even when the meaning is figurative.

Given the complexity of metaphors in their comparative nature, their directionality, their incorporation of figurative language and their requirement for coherent meaning, their comprehension appears to constitute a challenging task that may differentiate gifted from non-gifted students in the area of analytical thinking in the verbal domain. In this paper, a subtest of Aurora-*a*, a set of assessments based on Sternberg's theory of successful intelligence primarily designed for the identification of gifted students (Ortony, 1979a; Ortony et al., 1985; Sternberg, 1999, 2005, 2010; Tourangeau & Sternberg, 1982; Verbrugge & McCarrell, 1977), is presented. This subtest, *Metaphors*, is intended to assess analytical thinking by requiring the student to establish a relationship between two disparate words (i.e. things, actions, concepts) by stating how they are alike, in a way that makes logical sense, even though the words may differ widely or present an unusual relationship. This study investigated: (1) whether metaphor comprehension constitutes an appropriate task for the identification of gifted students as compared to a commonly used measure of academically-related abilities; and further; (2) what metaphor characteristics may contribute to a better differentiation of gifted versus non-gifted performance. These questions were explored in two independent studies.

Study 1 focused on the overall performance of the *Metaphors* subtest in a student sample and questioned whether, based on student performance, a group of gifted students might be distinguished from non-gifted (as identified by traditional academic achievement tests). It was hypothesised that, regardless of item difficulty, gifted students would perform better than non-gifted across all items of the subtest. Study 2 aimed to better understand the items that best differentiated gifted from non-gifted students by focusing on two fundamental characteristics of metaphors that may significantly influence how they are comprehended: (1) similarity (how closely two things resemble each other, in character, quantity or appearance, without being identical); and (2) concreteness (the degree to which a thing may be ascertained by the senses, with 'abstract' signifying its opposite). Evidence from the literature on metaphors suggests that when the target and source terms are very similar, a categorisation strategy is often used to assist comprehension. When an unfamiliar or novel pairing of words is presented, the comparative process is engaged (and novel features may be discovered; Chiappe & Kennedy, 2001). In addition, abstract terms have been found to be more relational or schematic in nature, evoking other abstract concepts such as situations and emotions, suggesting that they are held in representational frameworks organised by associations (as opposed to categories for concrete terms; Crutch, 2006; Crutch & Warrington, 2005). Therefore, we hypothesised that less similar item-pairs and those involving more abstract targets may require a higher level of analytical thinking to process, and such item-pairs may thus better differentiate higher- from lower-ability (i.e. gifted from non-gifted) students.

Study 1: Does *Metaphors* distinguish gifted students?

Here, we explored the overall and item-level performance of *Metaphors* in a sample of 465 students, including a group of students previously identified as gifted using a standardised test of mathematical and reasoning abilities (Middle Years Information System, MidYIS; Centre for Evaluation and Monitoring, 2010). We made an assumption that those previously labelled would possess a higher level of analytical ability and hypothesised that gifted students would perform better than non-gifted across all items of the *Metaphors* subtest, regardless of items 'difficulty' (i.e. sample mean performance).

Method

Participants

Participants in this study were 465 students (50.5 per cent female) attending schools in the north-east of England from a culturally homogenous, primarily Caucasian population of low to mid SES (age ranged from 10.19 to 11.19; M=10.69; SD=0.28). The sample was ascertained in a region where all schools could be recruited and in which, because of social disadvantage, there was likely to be a significant number of children whose high potential abilities had not been identified.

Instruments

The MidYIS (Centre for Evaluation and Monitoring, 2010; Tymms & Coe, 2003) is a baseline assessment of developed ability and aptitude for learning designed for students entering secondary school (during Term 1 of Year 7 or at the end of Year 6). The MidYIS has four sections: vocabulary; mathematics; skills (proofreading and perceptual speed and accuracy); and non-verbal tasks (primarily visual spatial reasoning and logical thinking). Vocabulary and mathematics scores in particular are predictive of student academic achievement (Centre for Evaluation and Monitoring, 2010). Scored by computer, the MidYIS overall score was calculated by averaging the subtest scores for the vocabulary, mathematics, nonverbal, and skills sections.

Metaphors is comprised of nine open-ended response items, each of which presents a metaphorical relationship in the form of an explicit comparison (similes). Students are asked to formulate a coherent relationship between the source and target terms. Item example: 'Homework is like health food because…'. Example response: 'It is good for you even though you might not like it'. *Metaphors* responses are scored by trained raters on two scales: accuracy (task appropriateness; 0–2) and ability (analytical thinking; 0–4). Only the ability scores were used in this study. Low scores reflect comparisons that present simple physical features (e.g. 'they are both blue') or common membership in a broad category (e.g. 'they are both found outside') because they are most obvious and/or common similarities. Similarities that are more subtle, perhaps combining abstract concepts with concrete properties, and outlined specifically and elaborately are rated with higher scores. *Metaphors* total scores were calculated as the average of the child's ability score for each item. (For more on the Aurora Battery, see Chart et al., 2008; Kornilov et al., 2011; Sternberg, 2010.)

Procedure

The administered tests were being piloted for the purposes of district-level screening and, therefore, were undertaken as part of the schools' regular activities under the auspices of the Local Education Authority. The Aurora Battery (including *Metaphors*) was administered to all classes in multiple sessions over the course of several days, one session per day. All of the students were in Year 6/5th grade at the time that *Metaphors* was administered; they were finishing Y7 when they took the MidYIS, approximately one-and-a-half years later. In line with standard practice and previous work (Tymms & Coe, 2003), we used a 90th-percentile criterion to identify gifted students using the MidYIS overall score.

Results

Preliminary analyses

Inter-rater agreement on *Metaphors* scoring was estimated using a double coding procedure of 160 protocols. Kappa statistic, performed for each *Metaphors* item to determine absolute agreement in scoring (exact same score) among raters suggested substantial agreement (Kornilov et al., 2012; Tan et al., 2009), with Kappa values ranging from .65 (Item 4 ability) to .94 (Item 2 accuracy), with a median of .80 (p<.001, 95 per cent CI [.71, .91]). In addition, two-way mixed intraclass correlation coefficients (ICC) were computed to estimate the rank-order consistency of the raters' scoring, regardless of their differences in severity and discrimination. These coefficients revealed excellent inter-rater agreement with ICC ranging from .90 to .98 (median ICC=.94, p< 001). Given the quality of the *Metaphors* ratings, scores were computed by summing the item scores for each item. The internal consistency coefficients of the ability composite score were acceptable (Cronbach's alpha=.75).

Gifted vs. Non-gifted performance

Mean performance on the *Metaphors* subtest differed significantly between gifted (top 10 per cent on MidYIS test performance) and non-gifted students (F[1,463]=5.41, p<.05), with a higher mean level for the gifted (M=12.74; SD=5.86) versus non-gifted students (M=10.96; SD=5.07). No interaction effects according to age and gender accounted for the difference between gifted and non-gifted students. Next, a MANOVA was performed on the *Metaphors* item-level ability scores to determine which items may be more effective for discriminating gifted vs. non-gifted students. Table 1 presents the means and standard deviations obtained for each item by the participant sample as a group, as well as individually for each group (gifted and non-gifted). Figure 1 presents the z-score means for gifted and non-gifted groups.

As represented in Table 1, *Metaphors* items display varying levels of 'difficulty' (i.e. a range of overall mean scores); some items appear to differentiate gifted from non-gifted students more clearly, while others seem not to differentiate them at all (Figure 1).

Specifically, gifted/non-gifted group differences in performance for each item ranged from null difference (Δ_{item4}=–.07, p=.57) to moderate difference (e.g. Δ_{item3}=.54, p=.003), yet interestingly, the overall group performance (reflecting item difficulty) for these non-differentiating and differentiating items was rather similar (M_{item3}=1.68, M_{item4}=1.61). In all, four items showed significant differences between groups (Items 3, 5, 7 and 8) ranging from .33 to .54 points. These items do not correspond to the most difficult items for the group as a whole.

Discussion

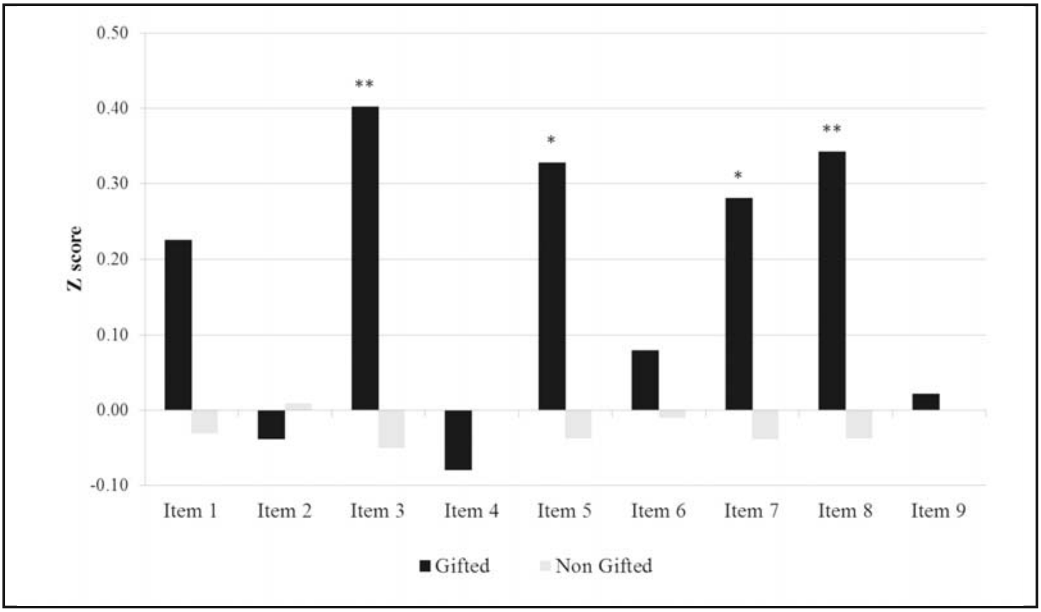
Metaphors appears to be an internally consistent subtest that provides a set of items that can be scored with a high level of inter-rater reliability and presents a range of item difficulties. Gifted students who perform exceptionally well on an assessment for academic aptitude and ability also perform significantly better overall on *Metaphors* than their non-gifted counterparts, providing some evidence for *Metaphors*' predictive validity. Hence, this subtest may be useful for the identification of gifted populations. Yet, interestingly, some items differentiate gifted from non-gifted better than others, and these items' ability to discriminate groups did not appear to be related to item difficulty. Therefore, these results suggest that item difficulty is not the essential criterion for the differentiation of gifted versus non-

Table 1: Descriptive statistics and differences between gifted and non-gifted group.

	N_{NG}	N_G	M_A	SD_A	η^2	Dif $M_G - M_{NG}$	M_G	SD_G	M_{NG}	SD_{NG}	F	*p*
All Items	415	50	11.130	5.200	.012	1.780	12.740	5.860	10.960	5.070	5.408	.012
Item 1	415	50	1.620	.978	.006	.250	1.840	1.010	1.590	.970	2.864	.091
Item 2	415	50	1.120	1.044	.000	–.050	1.080	1.010	1.130	1.050	.084	.772
Item 3	415	50	1.680	1.193	.019	.540	2.160	1.620	1.620	1.190	9.156	.003
Item 4	415	50	1.610	.880	.001	–.070	1.540	.880	1.610	.880	.319	.572
Item 5	415	50	.950	1.066	.013	.390	1.300	1.270	.910	1.030	6.015	.015
Item 6	415	50	1.360	1.009	.001	.090	1.440	1.060	1.350	1.000	.378	.539
Item 7	415	50	1.250	1.032	.010	.330	1.540	.940	1.210	1.030	4.610	.032
Item 8	415	50	.490	.788	.014	.300	.760	1.000	.460	.750	6.748	.01
Item 9	415	50	1.060	.914	.000	.020	1.080	.880	1.060	.920	.032	.858

Note: N_G=N, gifted'; N_{NG}=N, non-gifted; M_A=mean, across sample; SD_A=stdev, across sample; M_G=mean, gifted; M_{NG}=mean, non-gifted; SD_G=stdev, gifted; SD_{NG}=stdev, non-gifted.

Figure 1: *Z*-scores means for gifted and non-gifted groups.



Note: $^{**}p<.01$; $^{*}p<.05$

gifted students (at least, as identified using the MidYIS), and that other features of metaphor comprehension should be investigated to understand the source of group differences in *Metaphors*' item performance.

Study 2: Characterising the items of *Metaphors*

Study 2 investigated the nature of the items that best differentiated gifted from non-gifted students with regard to similarity and concreteness, two fundamental aspects of metaphor comprehension. Based on the literature (see Introduction) we hypothesised that less similar items-pairs and those involving more abstract targets may require a higher level of analytical thinking (or analytical ability) to process, and thus would better differentiate gifted from non-gifted students.

Method

Participants

Two independent samples of adults (hereafter labelled 'judges') were involved in this study. They were all adult members of the university community, representing students and a range of professions. All subjects reported English as their native language. Forty-three adults between 17 and 63 years of age (M=36.93, SD=13.60; 17 male, 26 female) filled out a 'concreteness survey'. Another sample of 36 adults ranging from 16 to 66 years of age (M=35.03, SD=14.31; 16 male, 20 female) filled out a 'similarity survey'.

Instruments and procedure

Based on survey instruments developed in previous studies, the source and target terms in *Metaphors* were rated on the concreteness of each term (Landis & Koch, 1977) and on the similarity between the target and source presented in each item (Gibb & Wales, 1990). In the concreteness survey, judges were provided with a brief written definition of concreteness ('the degree to which a thing can be ascertained by the senses'), then they rated each of the 18 *Metaphors* terms (i.e. the target and source terms of the nine *Metaphors* items) on a bipolar Likert-type scale ranging from '1=More concrete' to '5=More abstract'. Correspondingly, the similarity survey provided a brief definition of similarity ('having a resemblance in a general way, in character, quantity or appearance, without being identical'), and each pair of words (i.e. from the nine *Metaphors* items) was rated for the degree of similarity between the target and source terms on a Likert-type scale ranging from '1=Very dissimilar' to '5=Very similar'. Information on gender, age, and native English speaker status of the judges was collected on both surveys. For each survey, three alternative forms A, B and C, presenting the terms (concreteness survey) or items (similarity survey) in a random order were used to control for possible order effects.

Data analyses

The data obtained in this study were analysed using a Many-Facets Rasch Measurement approach (Chiappe et al., 2003; Johnson & Malgady, 1979) as implemented in FACETS (MFRM; Linacre, 1994). Item-level response probabilities were modelled separately for concreteness and similarity in a three-facet MFRM model with facets corresponding to judges, order (i.e. survey Form A, B or C), and items. For concreteness, 18 words were assessed by 43 judges, yielding a total of 774 observations; for similarity, nine items were assessed by 36 judges, yielding a total of 324 observations. The ratings for the individual items provided by the judges were used to reach a stable calibration of items with respect to the degree of similarity and the degree of concreteness on a common logit scale. These were exempted of differences in judges' severity, discrimination and, other systematic biases such as item order presentation, permitting their direct comparison.

Results

For the similarity data, the local fit indices (indicating the extent to which the data fit the measurement model) were examined to investigate the quality of the measurement model. Standard guidelines provided by Bond and Fox (Linacre, 2009) suggest that values falling within the range from .50 to 1.70 are deemed acceptable for measurement model construction. After removing five outlier judges with high degrees of misfit, an adjusted model revealed acceptable fit for the remaining judges and no other misfitting elements. This resulting Rasch model explained 39.09 per cent of the variance (global fit $\chi^2[234]=728.0054$, $p<.001$). The judges differed in their leniency of similarity ratings ($\chi^2[30]=87.9$, $p<.001$) while no item-order effects were found ($\chi^2[2]=5.2$, $p>.05$). As expected, significant differences across items were obtained for item similarity ($\chi^2[8]=33.1$, $p<.01$). We found high inter-rater reliability (.77), with an exact agreement of 20.9 per cent for ratings of similarity (23 per cent expected by the model).

Similarly, the first concreteness model explained 69.38 per cent of the variance ($\chi^2[708]=1519.4232$, $p<.001$). After eight misfitting judges were removed, the resulting model explained 79.39 per cent ($\chi^2[419]=756.4127$, $p<.001$). Results indicated judges' differences in item ratings ($\chi^2[34]=145.3$, $p<.001$), although high inter-judge reliability was observed (.78), with an exact agreement between judges being 48.0 per cent (47.4 per cent expected by the model). No item order effect was observed ($\chi^2[2]=2.7$, $p=.25$), while expected significant differences in item concreteness were obtained ($\chi^2[17]=775.7$, $p<.001$). Table 2 presents the resulting logit scores for the similarity between item-pairs and the concreteness of target and source terms within each item.

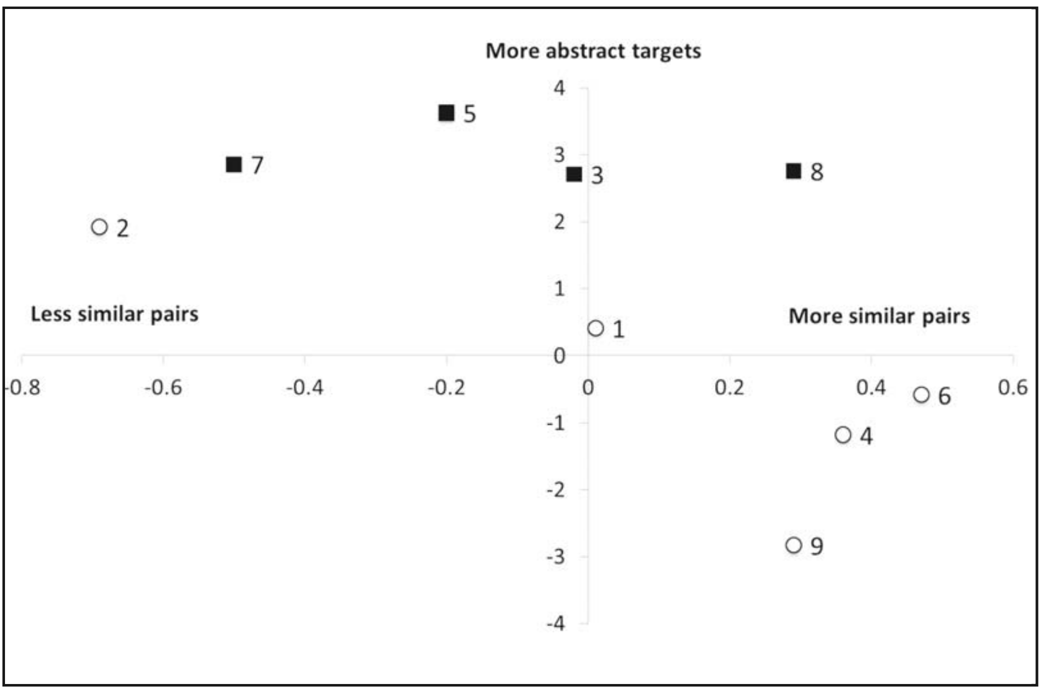
Logit scores for item-pair similarity ranged from –0.69 to 0.47 (range, 1.16 logits); from –2.83 to 3.63 (range, 6.46 logits) for target concreteness, and from –3.94 to 1.44 (range, 5.38 logits) for source concreteness. To combine these results in a single illustration of Metaphors' items features, we plotted the logit scores for similarity against the difference between each item-pairs' concreteness scores (Figure 2). Based on the calibrated scores we also computed the logit difference between

Table 2: Logit scores for item-pair similarity, target and source concreteness, and item-pair difference in concreteness.

Item #	Item-pair similarity	Target concreteness	Source concreteness	Item-pair difference in concreteness
1	0.01	0.41	-1.95	2.36
2	-0.69	1.92	1.44	0.48
3	-0.02	2.71	0.64	2.07
4	0.36	-1.18	-2.11	0.93
5	-0.2	3.63	-0.04	3.67
6	0.47	-0.58	-0.17	0.41
7	-0.5	2.86	-3.94	6.8
8	0.29	2.76	-1.46	4.22
9	0.29	-2.83	-2.11	0.72

Note: Less similar item-pairs and more concrete terms <0.

Figure 2: Item-pair similarity plotted against each item's target concreteness.



Note: Numbers indicate item numbers;
■ items that distinguish gifted from non-gifted;
○ items that do not distinguish gifted from non-gifted.

target and source concreteness as an indication of the relative distance between the concreteness of the item-term pairs, since a metaphor with a very concrete source and a very abstract target may present a qualitatively different problem from a metaphor in which the target and source are equally concrete.

Discussion

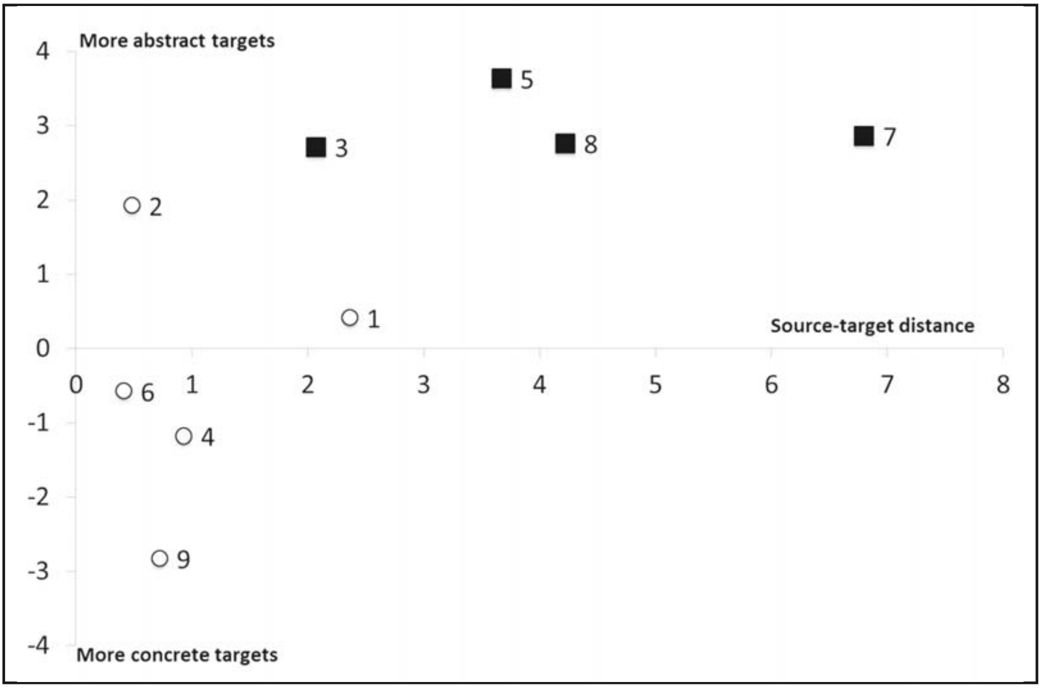
Study 2 revealed that beyond individual differences in item ratings (i.e. indicators of difficulty level), *Metaphors*' item-pairs could be reliably characterised according to their concreteness and similarity. A range of these characterisations were obtained for both aspects, yet the 18 terms covered a rather broad range of concreteness, while degrees of similarity remained within a much more circumscribed range. In all, *Metaphors*' nine items showed a range of similarity-concreteness combinations, with most items presenting more abstract targets in both more similar and less similar pairs (see Figure 2). Less abstract targets were found only in more similar pairs, possibly because concrete terms tend to be organised by categorisation (Crutch, 2006; Crutch & Warrington, 2005), facilitating the discernment of the superordinate category that contains both terms (2007).

To further understand why certain items better differentiated gifted from non-gifted students than others, the results of the two studies were combined. When item similarity is plotted against target concreteness, as in Figure 2, and we identify those items (3, 5, 7 and 8) that appeared to discriminate gifted students from non-gifted best (black squares), we see that they have generally more abstract targets and slightly less similarity between terms. This suggests that the hypothesis that gifted students are better able to map source to target meanings when more abstract meanings are involved is tenable. Also, item-pair similarity appears to contribute less to giftedness identification as discriminating items exhibit a range of item-pair similarity from less similar to more similar. However, when taking into account the distance between the concreteness of the source and target terms in an item-pair, results showed that the items that best discriminate gifted from non-gifted have the most abstract targets and tend to have greater differences in source-target concreteness. This is illustrated in Figure 3. Note that while these four items are clearly segregated according to target abstractness and source-target distance, their item difficulties vary (see Table 1, above), with item 3 being the easiest (total sample mean, *MA*=1.68) and item 8 the most difficult (*MA*=.49). This suggests that item difficulty is not the only aspect that contributes to the differentiation of gifted from non-gifted students when considering item-level performance.

General discussion

It is not at all unusual to target analytical thinking as an indicator of giftedness. There are several group-administered assessments that may be used to do this, most of them using a multiple-choice format. However, it is one of the goals of the Aurora Battery to move away from this single-format, single ability approach to gifted identification to explore gifted children's abilities in a more multi-dimensional fashion. Thus, the *Metaphors* subtest of Aurora addresses individual differences in the processing and comprehension of metaphors as an indicator of giftedness, primarily targeting analytical thinking in the verbal domain. Indeed, *Metaphors* behaved comparably against a standardised assessment generally used for the selection of gifted, with better performance by gifted students over non-gifted. In exploring this further, results of Study 2 further suggest that the understanding and manipulation of more abstract terms across semantic distance – that is, the ability to map the meaning of a concrete term onto that of a much more abstract term that is also quite dissimilar – may reflect the higher levels of analytic thinking of which gifted students are more capable. This is in line with previous studies on cognitive development and the

Figure 3: Target concreteness scores plotted against differences in target-source concreteness.



Note: Numbers indicate item numbers;
■ items that distinguish gifted from non-gifted;
○ items that do not distinguish gifted from non-gifted.

comprehension of metaphors and other linguistic complexities (Dryll, 2009; Fraser, 2003; Winner, 1997), whose results suggest a higher ability to discern semantic representations and categorisations in gifted children.

The weaknesses of our study include a fairly low number of items that cover the range of our chosen characteristics, concreteness and similarity. This might inform the next revision of the *Metaphors* subtest, as future studies on the topic should employ more items representing each of the item characteristic combinations (e.g. multiple items-pairs that have both a low level of similarity between source and target, and a low abstractness in the target term, as well as item-pairs that have the reverse 'profile'). In addition, it must be recognised that metaphors are complex entities, and therefore there are many more aspects of metaphor that may be examined with respect to their comprehension.

Despite these limitations, we have illustrated the potential of the *Metaphors* subtest to access analytical thinking for the differentiation of gifted students, and suggested important item characteristics that may help improve giftedness identification. A further benefit, we would argue, is that it may offer a richer picture of gifted students' thinking in general. *Metaphor* comprehension engages some aspects of creativity as it draws upon a person's ability to meaningfully associate two disparate items (Glucksberg 2008; Glucksberg & Keysar 1990). Yet, it has also been viewed as having pragmatic value for the facilitation of social relationships and the enhancement of communication (Tourangeau & Sternberg, 1982). Thus, overall, *Metaphors* may offer a more nuanced picture of a child's analytical skills. Accounting for these types of skills might constitute a better definition and assessment of giftedness that may be nurtured to fruitfulness both during and beyond formal education. Furthermore, understanding the 'texture' of these items will result in better guidance for designing new metaphor-based items for any assessment of analytical ability.

The authors

Mei Tan
Yale University, USA.
Baptiste Barbot
Yale University, USA.
Catalina Mourgues
Pontificia Universidad Catolica de Chile, Chile.
Elena L. Grigorenko
Yale University, USA.
Columbia University, USA.
Moscow State University, Russia.

Address for correspondence

Elena L. Grigorenko
Yale University,
230 South Frontage Road,
New Haven, CT 06519-1124.
Tel: 203.737.2316; Fax: 203.785.3002

References

Bond, T. & Fox, C.M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ, Lawrence Erlbaum Associates.

Centre for Evaluation and Monitoring (2010). *MidYIS assessments*. Accessed August 2012, from: http://www.midyis-project.org

Chart, H.E., Grigorenko, E.L. & Sternberg, R.J. (2008). Identification: The Aurora Battery. In J.A. Plucker & C. M. Callahan (Eds). *Critical issues and practices in gifted education* (pp.345–365). Waco, TX, Prufrock Press.

Chiappe, D.L. & Kennedy, J.M. (2001). Literal bases for *Metaphor* and *Simile*. *Metaphor & Symbol, 16*(3/4), 249–276.

Chiappe, D.L., Kennedy, J.M. & Smykowski, T. (2003). Reversibility, aptness, and the conventionality of *Metaphors* and *Similes*. *Metaphor & Symbol, 18*(2), 85–105.

Crutch, S.J. (2006). Qualitatively different semantic representations for abstract and concrete words: Further evidence from the semantic reading errors of deep dyslexic patients. *Neurocase, 12*(2), 91–97.

Crutch, S.J. & Warrington, E.K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain, 128*(3), 615–627.

Dryll, E. (2009). Changes in metaphor comprehension in children. *Polish Psychological Bulletin, 40*(4), 204–212.

Evans, V. (2010). Figurative language understanding in LCCM theory. *Cognitive Linguistics, 21*(4), 601–662.

Fraser, D.F.G. (2003). From the playful to the profound: What metaphors tell us about gifted children. *Roeper Review, 25*(4), 180.

Gentner, D. & Bowdle, B. (2008). Metaphor as structure-mapping. In R.W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought* (pp.109–128.). New York: Cambridge University Press.

Gentner, D. & Markman, A.B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45–56.

Gentner, D. & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development, 14*, 487–513.

Gibb, H. & Wales, R. (1990). Metaphor or Simile: Psychological determinants of the differential use of each sentence form. *Metaphor & Symbolic Activity, 5*(4), 199.

Gibbs, R.W. (Ed.) (2008). *The Cambridge Handbook of Metaphor and Thought*. New York: Cambridge University Press.

Gibbs, R.W. (2011). Evaluating conceptual metaphor theory. *Discourse Processes, 48*(8), 529–562.

Gibbs, R.W. & Tendahl, M. (2011). Coupling of metaphoric cognition and communication: A reply to Deirdre Wilson. *Intercultural Pragmatics, 8*(4), 601–609.

Glucksberg, S. (2008). How metaphors create categories – quickly. In R.W. Gibbs (Ed.), *The Cambridge Handbook of Metaphor and Thought.* New York: Cambridge University Press.

Glucksberg, S. & Keysar, B. (1990). Understanding metaphorical comparison: Beyond similarity. *Psychological Review, 97*(3–18).

Johnson, M.G. & Malgady, R.G. (1979). Some cognitive aspects of figurative language: Association and metaphor. *Journal of Psycholinguistic Research, 8*(3), 249–265.

Kornilov, S.A., Tan, M., Elliott, J.G., Sternberg, R.J. & Grigorenko, E.L. (2011). Gifted identification with Aurora: Widening the spotlight. *Journal of Psychoeducational Assessment*, 1–17.

Kornilov, S.A., Tan, M., Elliott, J.G., Sternberg, R.J. & Grigorenko, E.L. (2012). Gifted identification with Aurora: Widening the spotlight. *Journal of Psychoeducational Assessment, 30*(1), 117–133.

Kovecses, Z. (2010). A new look at metaphorical creativity in cognitive linguistics. *Cognitive Linguistics, 21*(4), 663–697.

Lakoff, G. & Johnson, M.G. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174.

Linacre, J.M. (1994). *Many-facet Rasch measurement.* Chicago: MESA Press.

Linacre, J.M. (2009). *Facets: Rasch measurement model computer program (version 3. 65.0).* Chicago: Winsteps.

Ortony, A. (1979a). Beyond literal similarity. *Psychological Review, 86*, 161–180.

Ortony, A. (Ed.) (1979b). *The Cambridge Handbook of Metaphor and Thought.* New York: Cambridge University Press.

Ortony, A. (Ed). (1993). *The Cambridge Handbook of Metaphor and Thought.* New York: Cambridge University Press.

Ortony, A., Vondruska, R.J., Foss, M.A. & Jones, L. (1985). Salience, similes, and the asymmetry of similarity. *Journal of Memory and Language, 24*, 569–594.

Sternberg, R.J. (1999). The theory of successful intelligence. *Review of General Psychology, 3*(4), 292–316.

Sternberg, R.J. (2005). The theory of successful intelligence. *International Journal of Psychology, 39*(2), 189–202.

Sternberg, R.J. (2010). Assessment of gifted students for identification purposes: New techniques for a new millennium. *Learning and Individual Differences, 20*(4), 327–336.

Sternberg, R.J. & Nigro, G. (1983). Interaction and analogy in the comprehension and appreciation of metaphors. *Quarterly Journal of Experimental Psychology, 35A*, 17–38.

Tan, M., Aljughaiman, A., Elliott, J.G., Kornilov, S.A., Ferrando-Prieto, M., Bolden, D.S., Adams-Shearer, K., Chart, H., Newman, T., Jarvin, L., Sternberg, R.J. & Grigorenko, E.L. (2009). Considering language, culture, and cognitive abilities: The international translation and adaptation of the Aurora Assessment Battery. In E.L. Grigorenko (Ed.), *Multicultural psycho-educational assessment* (pp.443–468). New York: Springer.

Tourangeau, R. & Rips, L. (1991). Interpreting and evaluating metaphors. *Journal of Memory and Language, 30*, 452–472.

Tourangeau, R. & Sternberg, R.J. (1982). Understanding and appreciating metaphors. *Cognition, 11*(3), 203–244.

Tymms, P. & Coe, R. (2003). Celebration of success of distributed research with schools: The CEM Centre, Durham. *British Educational Research Journal, 29*(5), 639–653.

Verbrugge, R.R. & McCarrell, N.S. (1977). Metaphoric comprehension: Studies in reminding and resembling. *Cognitive Psychology, 9*, 494–533.

Winner, E. (1997). *The point of words: Children's understanding of metaphore and irony.* Cambridge, MA: Harvard University Press.