Assessing the attitudes of students toward school subjects with the semantic differential and interactive visual metaphors

Josef Kundrát,[1] Karel Rečka,[1] Karel Paulík,[1] František Baumgartner,[1] Marek Malůš,[1] Tereza Benešová,[2] Kristina Thielová[1]

[1] Department of Psychology, Faculty of Arts, University of Ostrava
[2] Faculty of Electrical Engineering and Computer Science, Technical University of Ostrava

Psychologie a její kontexty 13(2), 2022, 63–79
https://doi.org/10.15452/PsyX.2022.13.0011

Abstract This study aims to provide preliminary evidence of the concurrent and structural validity of a novel instrument to measure students' attitudes toward school subjects. This new tool uses interactive software and measures attitudes less directly by asking students to change different visual features of objects displayed in a virtual three-dimensional space, thereby expressing their attitudes metaphorically (popularity/liking through size, importance through size, difficulty through weight, and relationship with a teacher through temperature). The results of this novel method were compared and related to the semantic differential, an already established instrument that is frequently used in the context of school evaluation and focuses on connotative meanings. The study was carried out on a sample of 147 Czech primary school students who rated three school subjects: Czech, English, and mathematics. We used structural equation modeling to analyze the measurement structure of both instruments and the relationship between them. The results showed that a single dominant factor (popularity/liking) can explain the majority of the variance shared between items. However, residual correlations suggested that two other factors are at play, namely importance (measured by the visual metaphor of size) and difficulty (measured by the visual metaphor of weight). The results did

not support a three-factor structure of the semantic differential, which is proposed by the theory of the instrument, since only two semantic differential items appeared to measure something different from popularity/liking. Furthermore, the visual metaphor of distance was the best indicator of popularity/liking, and although other metaphors related to popularity/liking as well, they showed discriminant validity in relation to other metaphors, as evidenced by their pattern of residual correlations with the semantic differential items. In summary, the results support the concurrent validity of the novel instrument utilizing interactive metaphors and also the discriminant validity of individual metaphors but question the proposed factorial structure of the semantic differential.
Keywords distance metaphor, size metaphor, attitude, school subjects, semantic differential.

Assessment of attitudes in the school environment

The assessment of students' attitudes toward school subjects and the school itself is an integral part of the school evaluation process. They are often reported to be an important indicator of the effectiveness of teaching and learning (Garné et al., 2005), since they influence attitudes toward learning, extracurricular activities (Lee, 2016), and also the behavior of students (Weinholtz & Stritter, 2009). Given the amount of time students spend at school, the assessment of the attitudes of students toward school subjects, education, and school is of undeniable importance for educational theory and practice (Şeker, 2013).

Traditionally, attitudes assessment tools have been divided into two categories: direct methods are based on explicit verbal statements, while indirect methods are not, but instead use projective techniques, physiological measurement, or observation (Fishman et al., 2021; Sollár, 2019). Questionnaires and interviews are the most widely used direct methods (Reid, 2006). These procedures are easier to administer and score but are more susceptible to response bias due to social desirability, response styles, or similar effects (Dodou & de Winter, 2014). Indirect methods include, among others, projective techniques or psychophysical measures (Fishman et al., 2021). In a school setting, attitudes and other psychological variables are measured mainly with attitudinal scales that contain response options with seemingly equal intervals. A Thurston scale, or method of successive intervals, a Likert scale, or agree–disagree rating scale, and the Gutmann scale operate on this principle, where respondents systematically differentiate their opinions and choose from statements on a continuum from positive to negative (Gure, 2015; Sollár, 2019).

One of the widely used inventories to measure attitudes in school contexts is the semantic differential (Vašťatková & Chvál, 2010). Miovský (2006) classifies this instrument as a psychosemantic method that taps into the evaluation of different objects and their content similarity, thus providing data to create a distance matrix of individual concepts (Hahn & Heit, 2015). According to Osgood et al. (1957), who pioneered the method, the goal is to detect similarities and differences in connotations between the concepts under investigation, and this limits response bias to some extent. Pöschl (2011) argues that it is not obvious to the respondents how their answers will be handled, as the items of the instrument do not represent explicit evaluative questions or statements directed at the attitude object. Instead, the method allows one to compare the positions of objects on a set of items. Respondents indicate the "position" of an attitude object on a set of bipolar items (usually with a seven-point response scale) with both ends (poles) representing opposite adjectives. Three latent factors are supposed to underlie item responses: (1) evaluation (impression quality, degree of popularity/liking); (2) potency (strength of the concept); and (3) activity (potential to elicit movement or change). Therefore, the meaning of individual concepts can be captured in a three-dimensional semantic space, the dimensions of which can be determined by factor analysis. However, Pöschl (2011) suggested complexity as another factor that should be considered.

The number of items that are used for the semantic differential can vary. On the one hand, increasing their number provides the opportunity to assess the content of concepts in finer detail; on the other hand, it places increased demands on respondents in terms of time and verbal comprehension. Nevertheless, the method shows acceptable validity and reliability, especially for measuring cognitive and emotional aspects of attitudes (Pöschl, 2011).

Attitudinal diagnostics using interactive visual metaphors

The aim of our research team is to develop a novel instrument for the assessment of attitudes that is appropriate for the school environment and that stands at the boundary between direct and indirect assessment. Consequently, it should combine some of the advantages of direct measurement methods (e.g., ease of administration through interactive software) and indirect methods (the semiprojective nature of the diagnostic material) and be more user-friendly for both test administrators and respondents. Its theoretical background and practical application in a school context have been described in more detail in previous publications (see, e.g., Kundrát et al., 2017; Kundrát & Paulík, 2019; Kundrát & Rojková, 2021; Kundrát et al., 2022).

The main goal of the research project is to standardize and validate the instrument. The instrument asks students to evaluate school subjects through four visual metaphors: distance, size, temperature, and weight. To be more specific, students can change the parameters of an object (its distance, size, etc.) that is displayed in a three-dimensional virtual space, which is part of an interactive software environment. The instrument gradually evolved by utilizing feedback from respondents.

As mentioned above, the instrument includes four visual metaphors: distance, size, temperature, and weight. The distance metaphor is supposed to express the popularity/liking of a concept (i.e., a school subject) and is implemented in the following way. The name of the school subject (e.g., "Mathematics") is displayed in the three-dimensional space and students can move it closer or farther away. The idea of using visual distance as a metaphor to express liking is based on a close relationship between proximity and popularity/liking, or distance and unpopularity/disliking in various contexts. The relationship between psychological distance and evaluation was first utilized by Bogardus (1933) in his attitudinal scale of social distance. More recently, Williams et al. (2014) reported that object appraisal can be inferred from its placement. The study by Kundrát and Rojková (2021) reached similar conclusions: respondents placed negatively perceived concepts further away, while positively perceived concepts closer. Similarly, Marmolejo-Ramos et al. (2019) also used an interactive three-dimensional space and found that people placed positive concepts in the closest proximity, neutral ones somewhere in the middle, and negative ones farthest away.

Other metaphors that the instrument now includes have been gradually added as students expressed the need to evaluate more aspects of school subjects. Nevertheless, the emotional dimension of liking-disliking the subject, captured by the spatial proximity–distance metaphor, still represents the main evaluative dimension of the instrument. The choice of size as a metaphor to express importance was based on studies that relate the size of an object to how influential, valuable, and important it is perceived to be. For example, Josephs et al. (1994) and later Hasegawa (2020) concluded that, regarding physical dimensions, people in general prefer larger objects over smaller ones. This suggests that physical size may play an important role in the formation and assessment of preferences. The tendency to prefer larger objects was also demonstrated in a study by Silvera et al. (2002). Thus, the results are consistent with the idea that object size can serve as a heuristic cue for object judgment. For example, Meier (2008) argued that stimulus size is used as a heuristic cue for stimulus valence during the encoding process. In our instrument, respondents, after placing the object, can use a mouse scroll wheel to change the size of the object. In this way, they can express their perceived importance and meaning. The expression of popularity/liking through distance and importance through size proved to be a good match, as respondents who use our instrument generally find it easy to understand. However, students frequently communicated the need to address two additional aspects that they considered important in evaluating school subjects, namely their relationship with the teacher or the difficulty of the subject. Therefore, based on this feedback and literature review, we extended the instrument by adding the metaphors of temperature and weight.

The investigation of the "warmth as affection" metaphor has its roots in the study of Asch (1946), who found warmth to be one of the key aspects in judgement of other people. More recently, Ijzerman and Semin (2009) investigated whether social closeness depends on perceived temperature and found that participants who were placed in a warmer environment experienced a greater sense of social closeness than participants who were put in a colder environment. Therefore, this research supports the notion that temperature affects social closeness. According to Williams and Bargh (2008), a brief physical experience of cold or warmth is sufficient to alter interpersonal perceptions. As predicted by the authors, cold exposure decreased social warmth, while warm exposure increased it.

The metaphor of temperature was chosen to express the relationship with the teacher (of the school subject in question). In the interactive software environment, students can change the color of the concept using a bipolar color scale that smoothly transitions from a cold blue to a warm red. In this way, students can metaphorically express how "cold" or "warm" they perceive their relationship with the teacher to be.

Finally, the metaphor of weight was added to the instrument. Students commonly use the metaphor of weight to describe how difficult the subject is, since in Czech, if something is difficult or hard, it is described as "heavy", and if something is easy and effortless, it is described as "light". Therefore, we added a "weigh scale" to our instrument. Similar to the temperature, it is a bipolar color scale, but uses a different color scheme, gradually transitioning from lighter to darker gray. Students are instructed to express an easier object in a lighter shade and a more difficult object in a darker shade of gray.

Similarities and differences between semantic differential and interactive metaphors

To test the concurrent validity of our instrument that uses interactive metaphors, we compared its results with the semantic differential. We chose the semantic differential because it shares certain properties with our instrument. The semantic differential adapted by Pöschl (2011) is intended for the same population (i.e., Czech primary school students). Furthermore, it is one of the few standardized methods available in the Czech Republic to measure attitudes toward school subjects. The similarity also lies in the semidirect nature of the questioning, as neither instrument uses Likert-type, agree-disagree items, nor asks explicit questions about attitudes. Although our instrument was not designed to assess the connotations of concepts, but utilizes semiprojective features to express various aspects of attitudes, we assumed several relationships between interactive metaphors and the semantic differential. Specifically, we expected the strongest relationship to emerge between these pairs of variables: the metaphor of distance and the evaluation factor, the metaphor of size and the potency factor, the metaphor of weight and the complexity factor, and finally the metaphor of temperature and the activity factor. Our reasoning is as follows. The first pair (i.e., the metaphor of distance and the evaluation factor) represents the main evaluative "dimensions" (popularity/liking) in the respective instruments. For size, the importance metaphor, we expect a close relationship with the potency factor, indicating the power or strength of the concept. The complexity factor (or better, the complexity item, as this factor is represented by only one item) is closest in meaning to the concept's weight metaphor, and therefore we expect a strong relationship here.

Finally, the activity factor is defined as "the degree of energy that produces a change or movement" (Pöschl, 2011, p. 8). The expected correlation with the metaphor of temperature, which is supposed to be the quality of the relationship with the teacher, is based on personal experience that a good relationship with the teacher is related to increased attention and engagement of students. However, we admit that a priori support for this hypothesis is weaker.

However, due to the differences between the two methods, our research questions are of a more general and exploratory nature. In summary, we had the following research questions: How are the factors of the semantic differential and visual metaphor related? Can the responses of students to the semantic differential items and visual metaphors be explained by a single latent variable (general attitude, or subject popularity/liking)?

Research sample

The selection of students was deliberate, without using probabilistic sampling, as our preliminary pilot study found that many respondents experienced difficulties understanding the semantic differential items. Therefore, eligible students were selected based on information from teachers. Specifically, we asked teachers to select students with verbal comprehension sufficient to understand the semantic differential items without difficulties. We comment on this decision and its implications in the discussion section.

Participation was voluntary. Informed consent was obtained from all participating students and all participants were informed that they could withdraw their consent at any time. Additionally, we also obtained informed consent from the parents or other legal guardians of the students who agreed to participate. The data were anonymized before being exported from the software database and analyzed.

The research sample consisted of 147 students from three Czech primary schools. The students rated three school subjects: English, mathematics, and Czech language. Therefore, the total number of observations (i.e., the number of ratings of all three subjects from all students) was 147 × 3 = 441. The number of boys (n = 76, 52%) and girls (n = 71, 48%) was almost the same. There were 59 (40%) 7th grade students, 46 (31%) 8th grade students, and 42 (29%) 9th grade students. The composition of the sample in terms of both gender and grade is shown in Table 1.

Table 1

Composition of the research sample in terms of grade and gender

Gender	7. grade			8. grade			9. grade		
	n	$\%_{col}$	$\%_{row}$	n	$\%_{col}$	$\%_{row}$	n	$\%_{col}$	$\%_{row}$
Boy	29	49	38	24	52	32	23	55	30
Girl	30	51	42	22	48	31	19	45	27

Note: $\%_{col}$ – column relative frequency (in percent), $\%_{row}$ – row relative frequency (in percent).

Statistical data analysis

Statistical data analysis was performed in R version 4.2.1 using the following packages: *dplyr* (version 1.0.10; Wickham et al., 2022) for data transformation, *ggplot2* (version 3.3.6; Wickham, 2009) for plotting, *psych* (version 2.2.5; Revelle, 2017) for computing descriptive statistics and reliability estimates, *lavaan* (version 0.6-12; Rosseel, 2012) for estimating structural models, and *tidySEM* (version 0.2.3; van Lissa, 2022) for creating structural diagrams.

First, we calculate basic descriptive statistics. Then, we estimate measurement models for the semantic differential items using the *lavaan* package. We test both a simple unidimensional model and a three-factor model, advocated by the author of the semantic differential (Pöschl, 2011). For visual metaphors, we also evaluate the fit of the unifactorial model. Finally, we create a joint model with two latent variables, measured by the semantic differential items and visual metaphors, respectively. To estimate the models, we use a robust maximum likelihood estimator (*MLR*), which is suitable when indicators (items) are not normally distributed or are more ordinal than continuous in nature (Li, 2015). As fit indices, we report the χ^2-test of the model, the comparative fit index (*CFI*), the root mean square error of approximation (*RMSEA*), and the standardized root mean square residual (*SRMR*). The χ^2-test is a test of the exact fit of the model to the data and is almost always significant for larger samples (more than 200 cases). The CFI is a measure of the improvement in fit compared to a "null model" (which assumes that the covariances between variables are zero). A higher CFI value indicates a better fit to the data, with values of 0.90 or more generally considered acceptable. RMSEA is a measure of how far the proposed model differs from a model that would fit the data perfectly; therefore, lower values indicate a better fit to the data, with values of 0.08 and below usually considered acceptable. The SRMR is a standardized measure of how much the correlations predicted by the model differ from those observed, but unlike the RMSEA or the CFI, the SRMR does not include any penalty for model complexity. Lower values indicate a better fit to the data and values of 0.08 and below are usually considered acceptable. For a review of fit indices and a discussion of their cut-offs, see, e.g., Kenny (2020) or Xia and Yang (2019). Furthermore, we also analyze the differences between the ratings of different school subjects and report bivariate correlations between the semantic differential items and visual metaphors. During the analysis, we use $\alpha = 0.05$ as the significance level and report 95% confidence intervals for effect sizes.

Results

Table 2
Descriptive statistics of quantitative variables

Variables	*M*	*SD*	*Rng*	*Min*	*Q1*	*Mdn*	Q3	Max	Skew	Kurt
Semantic differential										
Evaluation factor										
1. Useless/useful	6.07	1.09	1–7	1.00	6.00	6.00	7.00	7.00	−1.44	2.53
2. Uniform/diverse	5.06	1.47	1–7	1.00	4.00	5.00	6.00	7.00	−0.63	0.07
3. Ugly/beautiful	4.57	1.65	1–7	1.00	4.00	4.00	6.00	7.00	−0.19	−0.70
4. Boring/interesting	4.45	1.79	1–7	1.00	3.00	5.00	6.00	7.00	−0.33	−0.85
Activity factor										
5. Slow/Fast	4.28	1.54	1–7	1.00	3.00	4.00	5.00	7.00	−0.31	−0.36
6. Old/Young	4.02	1.63	1–7	1.00	3.00	4.00	5.00	7.00	−0.06	−0.70
7. Passive/active	4.83	1.51	1–7	1.00	4.00	5.00	6.00	7.00	−0.64	0.04
8. Rigid/flexible	4.45	1.46	1–7	1.00	3.00	5.00	5.00	7.00	−0.27	−0.48
Potency factor										
9. Weak/strong	4.90	1.35	1–7	1.00	4.00	5.00	6.00	7.00	−0.50	0.13
10. Distant/close	4.92	1.65	1–7	1.00	4.00	5.00	6.00	7.00	−0.68	−0.35
11. Superficial/deep	4.65	1.38	1–7	1.00	4.00	5.00	6.00	7.00	−0.35	0.03
12. Narrow/wide	5.00	1.33	1–7	1.00	4.00	5.00	6.00	7.00	−0.54	0.23
Additional item										
13. Easy/difficult	4.23	1.78	1–7	1.00	3.00	5.00	6.00	7.00	−0.29	−1.03
Visual metaphors										
Distance	0.23	0.24	0–1	0.00	0.06	0.16	0.31	1.00	1.53	2.02
Size	0.52	0.28	0–1	0.00	0.30	0.44	0.71	1.00	0.51	−0.90
Temperature	0.69	0.26	0–1	0.00	0.56	0.73	0.89	1.00	−0.92	0.37
Weight	0.47	0.27	0–1	0.00	0.26	0.48	0.68	1.00	0.02	−0.88

Note: N = 441. *Rng* – theoretical range of values. *Q1* – 1st quartile (25th percentile), *Mdn* – median, *Q3* – 3rd quartile (75th percentile), Skew – skewness, Kurt – kurtosis.

Table 2 shows basic descriptive statistics of quantitative variables, including means, standard deviations, quartiles, skewness, and kurtosis. As can be seen, the semantic differential items were mostly only slightly skewed, so their means were higher than the midpoint of the response scale (four). Since higher values indicate more positive attitudes (gravitating toward the "positive" adjective) for all items, this means that students generally expressed more positive rather than negative attitudes toward the items. This tendency was most apparent for perceived usefulness, where the middle 50 % of the observed values ranged from 6 to 7, although the response scale had possible a range of values from 1 to 7. However, for all semantic differential items, the full range of the scale was used because the observed minimum and maximum values corresponded to the theoretical range of the response scale.

Regarding visual metaphors (see again Table 2), the students also used the full range of the response scale (0–1). In general, respondents tended to place objects closer relative to the center of the visual field, as the middle 50% of the distance values ranged from 0.06 to 0.31, so the distribution of these values was positively skewed. In the case of temperature, the students perceived their relationship with the teacher as "warmer" rather than "colder", because the middle 50% of the temperature values ranged from 0.56 to 0.73 and the distribution was positively skewed. Finally, for size and temperature, the distribution was approximately symmetric around the midpoint of the response scale (0.50).

Next, we analyzed the factor structure of the semantic differential. First, we tested the most parsimonious model, i.e., a single-factor (unidimensional) model. This model already showed a fit that could be considered acceptable or good, $\chi^2(64, N = 441) = 203.77, p < .001$, CFI = .932, RMSEA = .070 (90% CI [.061, .080]), SRMR = .046, although the authors assumed a three-factor structure, with the evaluation, activity and potency factors (see Table 2). The three-factor model showed a similar fit to the data as the single-factor model, $\chi^2(60, N = 441) = 196.91, p < .001$, CFI = .933, RMSEA = .072 (90% CI [.062, 0.082]), SRMR = .046. The difference in fit between the two models was not significant, $\Delta\chi^2(4, N = 441) = 7.60, p = .107$. Furthermore, all three factors were strongly correlated with each other ($r > 0.90$), suggesting a lack of discriminant validity. In other words, empirically differentiating between the three factors would be difficult. Therefore, we decided to work with a single-factor solution for the semantic differential, where the latent variable, or factor, represents popularity/liking of a school subject.

Then, we extended this model with another latent variable, measured by visual metaphors, since we assumed that a single common factor (also popularity/liking of a school subject) would also be a dominant source of variance. This extended model showed a worse fit to the data, $\chi^2(117, N = 441) = 459.93, p < .001$, CFI = .870, RMSEA = .082 (90% CI [.075, .089]), SRMR = .062. We inspected residuals and modification indices and found that the two residual correlations impaired the model fit most: between the *simple/complex* semantic differential item and the visual metaphor of weight, and between the useless/useful semantic differential item and the visual metaphor of size correlations. We allowed both of these residual correlations to be freely estimated, because we judged them to be theoretically justified. We believe that the first pair of items reflect perceived "difficulty" and the second pair perceived "importance", and therefore they share additional variance not explained by the dominant factor of popularity/liking. Allowing these two residual correlations resulted in a model with an acceptable fit to the data, $\chi^2(115, N = 441) = 324.24$, $p < .001$, CFI = .921, RMSEA = .064 (90% CI [.057, .072]), SRMR = .048, and this improvement in fit was statistically significant, $\Delta\chi^2(2, N = 441) = 118.92$, $p < .107$.

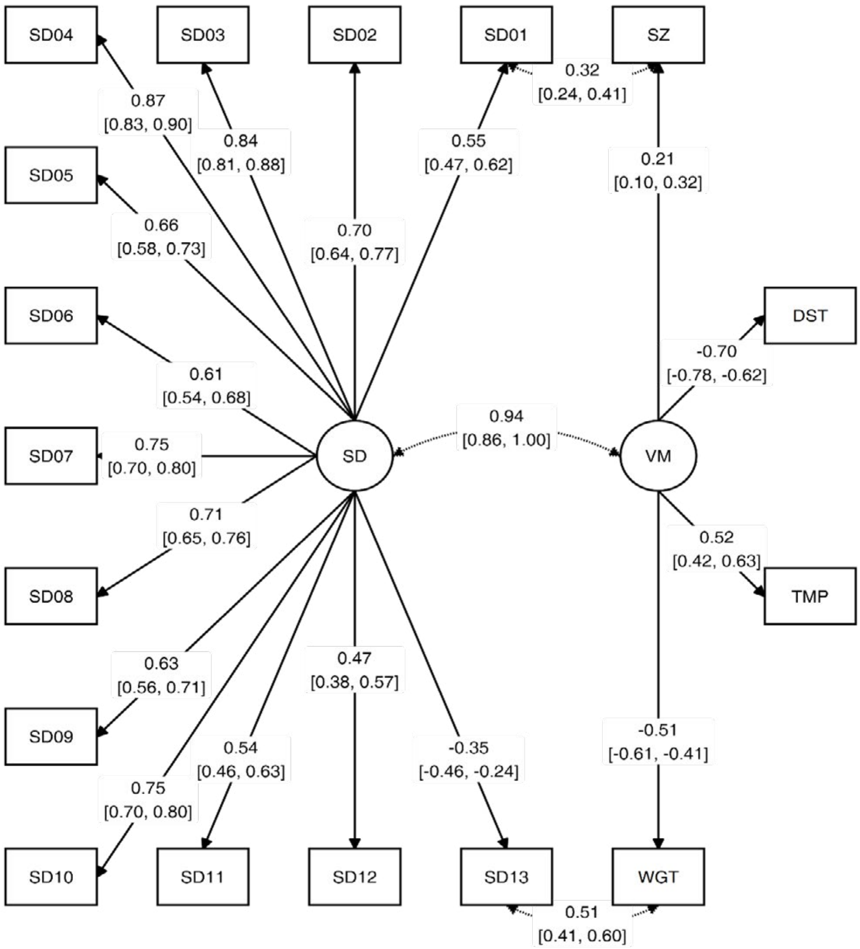
Figure 1 shows the structure of the final model, including standardized coefficients with 95% confidence intervals. As you can see, the two latent variables were very strongly correlated, $r = 0.94$, 95% CI [.86, 1.00]. Thus, they seem to be empirically indistinguishable from each other and both probably represent the overall popularity/liking of the school subject. Regarding visual metaphors, the distance was the best indicator of the respective latent variable and size the worst, since distance showed the highest factor loading ($\lambda = -0.70$, 95% CI [−0.78, −0.62]) while size showed the lowest one ($\lambda = 0.21$, 95% CI [0.10, 0.32]). Considering the semantic differential, the best indicators were more expressive items or items with stronger evaluative connotations, for example, *boring/interesting* ($\lambda = 0.87$, 95% CI [0.83; 0.90]) or *ugly/beautiful* ($\lambda = 0.84$, 95% CI [0.81, 0.88]).

In addition, we compared how students perceived the school subjects. Therefore, we calculated three total (composite) scores, since the previous analysis showed that three latent variables can underlie responses: popularity/liking, importance, and difficulty. Because the range of possible values for the semantic differential items (1–7) and the visual metaphors (0–1) were very vastly different, we first transformed the semantic differential items to the scale of 0–1 scale and then computed three new variables: popularity/liking, importance, and difficulty. Popularity/liking was calculated as the average of items 2–12 of the semantic differential and the visual metaphors of distance and temperature; the reliability estimate of this new variable was $\alpha = .91$. Importance was computed as the average of the first item of the semantic differential (*useful/useless*) and the visual metaphor of size; the reliability estimate for this new variable was $\alpha = 0.55$. Finally, difficulty was computed as the average of item 13 of the semantic differential (*easy/difficult*) and the visual metaphor of weight; the reliability estimate for this new variable was $\alpha = 0.74$. The correlation between popularity/liking and importance was $r = .45$ (95% CI [.37, .52]), $p < .001$; between popularity and difficulty $r = -.55$ (95% CI [−.55; −.40]); and between importance and difficulty $r = -.16$ (95% CI [−0.25, −0.07]), $p < .0011$.

Figure 2 shows the comparison of school subjects in the composite variables. As you can see, students rated English as more popular/likable than Czech and mathematics. However, all means were higher than the midpoint of the scale (0.50), so the students tend to perceive the school subjects as rather popular/likable than unpopular/dislikable. Furthermore, the students perceived English as the most important, mathematics as less important than English but more important than Czech, and Czech as the least important. Even in this case, the students tended to perceive the school subjects as important rather than unimportant, since all means were again higher than 0.50. As for difficulty, they perceived Czech and mathematics as more difficult than English. The means for Czech and mathematics were above 0.50, that is, closer to the "difficult" end of the scale. However, the mean for English was below 0.50, which is closer to the "easy" end of the scale.

Figure 1

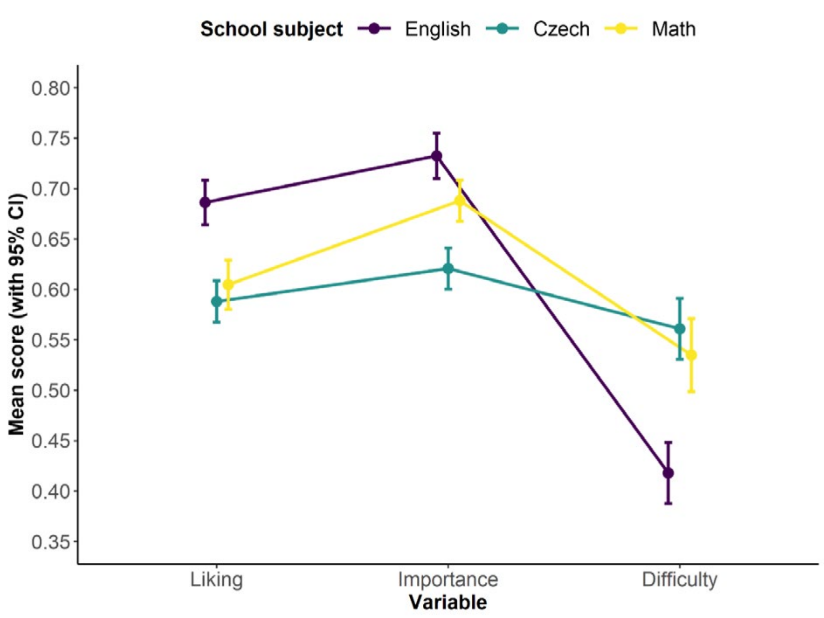
Path diagram of the final model, standardized solution



Note: The edge labels show standardized coefficients with 95% confidence intervals (in square brackets). All coefficients are significant at $p < .001$. Abbreviations: SD – semantic differential, SD01 – useless/useful, SD02 – uniform/diverse, SD03 – ugly/beautiful, SD04 – boring/interesting, SD05 – slow/fast, SD06 – old/young, SD07 – passive/active, SD08 – rigid/flexible, SD09 – weak/strong, SD10 – distant/close, SD11 – superficial/deep, SD12 – narrow/wide, SD13 – easy/difficult, VM – visual metaphors, DST – distance, SZ – size, TMP – temperature, WGT – weight.

Figure 2

Comparison of subjects in terms of perceived popularity, importance, and difficulty



Finally, we also report bivariate correlations (Spearman) between the factors and items of the semantic differential and the visual metaphors in Table 3.

Table 3

Spearman correlations between the semantic differential and visual metaphors

Semantic differential	Distance	Size	Temperature	Weight
Evaluation factor	−.66*** [−.74, −.56]	.32*** [.18, .45]	.46*** [.33, .58]	−.43*** [−.54, −.29]
1. Useless/useful	−.45*** [−.56, −.32]	.38*** [.24, .50]	.30*** [.16, .43]	−.25*** [−.38, −.11]
2. Uniform/varied	−.48*** [−.59, −.35]	.27*** [.14, .40]	.34*** [.20, .47]	−.25*** [−.38, −.11]
3. Ugly/colourful	−.62*** [−.71, −.51]	.17** [.04, .30]	.43*** [.30, .55]	−.49*** [−.60, −.36]
4. Boring/interesting	−.64*** [−.72, −.53]	.27*** [.13, .40]	.44*** [.30, .56]	−.40*** [−.52, −.27]
Activity factor	−.55*** [−.65, −.43]	.18** [.04, .31]	.41*** [.28, .53]	−.38*** [−.50, −.24]
5. Slow/Fast	−.40*** [−.52, −.27]	.12* [.01, .22]	.27*** [.14, .40]	−.22*** [−.35, −.08]
6. Old/Young	−.40*** [−.52, −.26]	.16** [.03, .29]	.30*** [.16, .43]	−.28*** [−.41, −.14]
7. Passive/active	−.52*** [−.63, −.40]	.13* [.01, .24]	.35*** [.21, .47]	−.36*** [−.48, −.22]
8. Rigid/flexible	−.44*** [−.56, −.31]	.13* [.01, .24]	.39*** [.25, .51]	−.32*** [−.45, −.18]
Potency factor	−.56*** [−.66, −.44]	.22*** [.08, .35]	.37*** [.23, .50]	−.47*** [−.58, −.34]

9. Weak/strong	−.45*** [−.57, −.32]	.17** [.04, .30]	.31*** [.17, .44]	−.31*** [−.43, −.17]
10. Distant/close	−.62*** [−.70, −.51]	.18** [.04, .31]	.35*** [.22, .48]	−.51*** [−.62, −.38]
11. Superficial/complex	−.37*** [−.49, −.23]	.15** [.03, .28]	.31*** [.17, .44]	−.34*** [−.46, −.20]
12. Narrow/wide	−.28*** [−.41, −.14]	.16** [.03, .29]	.15* [.02, .27]	−.24*** [−.37, −.11]
13. Easy/difficult	.41*** [.27, .53]	−.01 [−.11, .08]	−.15* [−.27, −.03]	.58*** [.47, .68]

Note: N = 441. Confidence intervals are in square brackets. Confidence intervals and p-values are corrected for multiple testing (Holm correction).
*p < .05; **p < .01, ***p < .001.

Discussion

In the present study, we compared two methods that purport to measure the attitudes of students, the established and still used semantic differential, adapted by Pöschl (2011), and a novel instrument that uses interactive metaphors. The two main latent factors that were identified, one explaining the variance in the semantic differential items and the other in visual metaphors, were almost perfectly correlated with each other, supporting the concurrent validity of the interactive metaphors. However, considering the factor structure of the semantic differential, we found that the three factors assumed by the author were not empirically supported. The three-factor solution did not show a significantly better fit to the data. Furthermore, all the factors in the three-factor solution were strongly correlated with each other, undermining their discriminant validity. However, the three-factor solution of the semantic differential method is not universally accepted. Takahashi et al. (2016) work with a two-dimensional model in the context of the study of mind perception. Similarly, a model with two dimensions is presented by Milin and Zdravkovic (2013).

Since the presented study involves a relatively small sample of respondents, it is difficult to evaluate the generalizability of the results to a broader population. Furthermore, we did not use all the possible scores the semantic differential can provide (e.g., analyzing the relative distance of school subjects and the self-concept in the semantic field) because it was beyond the scope of our research. Instead, we focused on the relationships between the semantic differential and visual metaphors. At the bivariate level, we identify the strongest correlation between the evaluation factor and distance, which is consistent with our assumption that these variables are the main carriers of the evaluative connotation in both methods. This is supported by the fact that they showed the strongest factor loadings in the final structural model, suggesting that they are the best indicators of popularity/liking.

As was mentioned, the semantic differential assumes that there are three latent variables (factors) that underlie item responses, and each factor is measured by four items. However, the data did not support this factor structure, but suggested that one factor is sufficient to explain the data. Only items *useless/useful* and *easy/difficult* seemed to measure something different from popularity/liking, namely perceived importance and difficulty of school subjects, since these items showed weaker factor loadings and significant residual correlations with some visual metaphors (the *useless/useful* item with the size metaphor and the *easy/difficult* item with the weight metaphor). Both latent variables (one measured by the semantic differential items and the other by visual metaphors) were almost perfectly correlated with each other; therefore, we believe that they are essentially the same variable, namely popularity/liking. As for visual metaphors, distance showed the highest factor loading, which was expected, since it is supposed to be a relatively pure indicator of popularity/liking. Weight and temperature had weaker, but still moderate and significant factor loadings. However, this result can be easily explained. Although weight is intended to primarily reflect perceived difficulty and temperature the relationship with the teacher, they are related to subject popularity/liking, although the direction of the effect cannot be determined without longitudinal data. For example, if a student excels in a subject, it is likely that they will like it, but it is also possible that liking a subject leads to increased engagement and effort, which supports the acquisition of subject knowledge and skills and results in a lower perceived difficulty. Similarly, it is possible that attitudes toward the teacher affect the attitudes toward the subject itself (e.g., "I don't like math because our teacher is boring"), or conversely, attitudes toward the subject might affect the attitudes toward the teacher (e.g., "I have a good relationship with our math teacher because I like math.").

The setup of the final study, its methodology, and sample selection were informed by a pilot study of both instruments in a small group of students (the data from the pilot study are not included in the current sample). The instruments were administered to seventh- and eighth- graders from two primary schools. Multiple students did not understand the meaning of some items (especially *uniform/varied* and *rigid/flexible*) and requested additional clarification. Although the questionnaire was administered following the standard instructions listed in the manual, many students complained that they simply did not fully understand the meaning of some items included in the semantic differential. As a result, the response rate was low. Based on this experience, we decided to deliberately include only students with sufficient verbal ability, nominated by teachers. In other words, the teachers selected students with sufficient vocabulary and verbal comprehension, and these were considered eligible to participate. Ultimately, all participants in the final sample rated both methods as easy to understand and did not need to ask any additional questions during the administration. This method of selection might have affected primarily the mean ratings of subjects, including popularity/liking, importance, and difficulty, since the results are more likely to reflect the attitudes of more gifted or motivated students, who generally perceive education, learning, and teachers more positively.

Despite its limitations, we believe that the presented study provides novel insights into the semantic differential and provides preliminary evidence supporting the validity of our novel instrument. Furthermore, a lack of evidence supporting the proposed three-

-factor structure of the semantic differential and its demands on verbal comprehension might provide an impetus to re-evaluate the psychometric properties and usability of this instrument in the current generation of students.

Source of funding

This contribution was financially supported and is dedicated to the TAČR Éta project TL03000240.

References

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology, 41*(3), 258–290. https://doi.org/10.1037/h0055756

Bogardus, E. (1933). A social distance scale. *Sociolgy and Social Research, 17,* 265–271.

Dodou, D., & de Winter, J. C. F. (2014). Social desirability is the same in offline, online, and paper surveys: A meta-analysis. *Computers in Human Behavior, 36*, 487–495. https://doi.org/10.1016/j.chb.2014.04.005

Fishman, J., Yang, C., & Mandell, D. (2021). Attitude theory and measurement in implementation science: a secondary review of empirical studies and opportunities for advancement. *Implementation Science, 16*(1), 87. https://doi.org/10.1186/s13012-021-01153-9

Garné, R. M., Wager, W. W., Golas, K. C., & Keller, J. M. (2005). *Principles of Instructional Design.* Wadsworth.

Gure, G. S. (2015). Different scale construction approaches used to attitude measurement in social science research. *International Journal of Research in Economics and Social Sciences, 5*(1), 26–44.

Hahn, U., & Heit, E. (2015). Semantic Similarity, Cognitive Psychology. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 579–584). Elsevier.

Hasegawa, K. (2020). The Size-Value Compatibility Effect. *Scientific Reports, 10*(1), 5383. https://doi.org/10.1038/s41598-020-62419-8

Ijzerman, H., & Semin, G. R. (2009). The thermometer of social relations: Mapping social proximity on temperature. *Psychological Science, 20*(10), 1 214–1 220. https://doi.org/10.1111/j.1467-9280.2009.02434.x

Josephs, R. A., Giesler, R. B., & Silvera, D. H. (1994). Judgment by quantity. *Journal of Experimental Psychology: General, 123*(1), 21–32. https://doi.org/10.1037/0096-3445.123.1.21

Kenny, D. A. (2020). *Measuring model fit*. http://www.davidakenny.net/cm/fit.htm#RMSEA

Kundrát, J., Cakirpaloglu, P., Rojková, Z., & Paulík, K. (2017). Psychologická vzdálenost jako nástroj zjišťování postojů u žáků. In. E. Maierová, L. Viktorová, J. Suchá, & M. Dolejš (Eds.), *PhD EXISTENCE 2017: Česko-slovenská psychologická konference (nejen) pro doktorandy a o doktorandech* (pp. 186–193). Univerzita Palackého v Olomouci.

Kundrát, J., & Paulík, K. (2019). Vztah psychologické vzdálenosti a postojového hodnocení. In I. Pitterová, & J. Výrost (Eds.), *Sociálne procesy a osobnosť 2018: Zborník príspevkov z medzinárodnej konferencie* (pp. 209–215). Spoločenskovedný ústav CSPV SAV.

Kundrát, J., & Rojková, Z. (2021). Psychological distance as a means of evaluation. *New Ideas in Psychology, 63*, 100900. https://doi.org/10.1016/j.newideapsych.2021.100900

Kundrát, J., Fabián, T., Skanderová, L., Beranová, H., Prachařová, D., Thielová, K., Husáková, L., Pavlíčková, A., Trochtová, V., & Peigerová, H. (2022). Zjišťování vztahů v malé sociální skupině prostřednictvím diagnostického software. In E. Aigelová, L. Viktorová, & M. Dolejš (Eds.), *PhD existence 12 „Změna": Česko-slovenská psychologická konference (nejen) pro doktorandy a o doktorandech* (pp. 488–498). Univerzita Palackého v Olomouci.

Lee, J. (2016). Attitude toward school does not predict academic achievement. *Learning and Individual Differences, 52*, 1–9. https://doi.org/10.1016/j.lindif.2016.09.009

Li, CH-H. (2015). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Marmolejo-Ramos, F., Arshamian, A., Tirado, C., Ospina, R. , & Larsson, M. (2019). The Allocation of Valenced Percepts Onto 3D Space. *Frontiers In Psychology*, *10,* 1–9. https://doi.org/10.3389/fpsyg.2019.00352

Milin, P., & Zdravkovic, S. (2013). Bi-dimensional semantic scales: the embodied maps of meanings. *Universitas Psychologica*, *12*(5), 1547–1562. https://doi.org/10.11144/Javeriana.UPSY12-5.bdss

Miovský, M. (2006). *Kvalitativní přístup a metody v psychologickém výzkumu*. Grada.

Meier, B. P., Robinson, M. D., & Caven, A. J. (2008). Why a Big Mac Is a Good Mac: Associations between Affect and Size. *Basic and Applied Social Psychology*, *30*(1), 46–55. https://doi.org/10.1080/01973530701866516

Osgood, CH. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. University of Illinois Press.

Pöschl, R. (2011). *Postoje žáků ke škole: dotazník pro žáky.* Národní ústav odborného vzdělávání.

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org

Reid, N. (2006). Thoughts on attitude measurement. *Research in Science & Technological Education, 24*(1), 3–27. https://doi.org/10.1080/02635140500485332

Revelle, W. (2017). *psych: Procedures for Psychological, Psychometric, and Personality Research.* Northwestern University. https://personality-project.org/r/psych/

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Şeker, H. (2013). In/out-of-school learning environment and SEM analyses on attitude towards school. In M. S. Khine (Ed.), *Application of Structural Equation Modeling in Educational Research and Practice* (pp. 135–167). Sense Publishers. https://doi.org/10.1007/978-94-6209-332-4_7

Sollár, T. (2019). Měření postojů. In J. Výrost, I. Slaměník, & E. Sollárová (Eds.), *Sociální psychologie: teorie, metody, aplikace* (pp. 341–351). Grada.

Silvera, D. H., Josephs, R. A., & Giesler, R. B. (2002). Bigger is better: The influence of physical size on aesthetic preference judgments. *Journal of Behavioral Decision Making*, *15*(3), 189–202. https://doi.org/10.1002/bdm.410

Takahashi, H., Ban, M. & Asada, M. (2016). Semantic differential scale method can reveal multi--dimensional aspects of mind perception. *Frontiers in Psychology*, 7, 1717. https://doi.org/10.3389/fpsyg2016.01717

Vašťatková, J., & Chvál, M. (2010). K využití sémantického diferenciálu při autoevaluaci školy. *Orbis Scholae, 4*(1), 111–128. https://doi.org/10.14712/23363177.2018.185

van Lissa, C. J. (2022). *tidySEM: Tidy structural equation modeling*. https://CRAN.R-project.org/package=tidySEM

Weinholtz, D., & Stritter, F. (2009). How to: Plan An Assessment of Student Attitudes. *Medical Teacher, 4*(3), 95–101. https://doi.org/10.3109/01421598209034759

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag.

Wickham, H., François, R. , Henry, L., & Müller, K. (2022). *dplyr: A grammar of data manipulation.* https://CRAN.R-project.org/package=dplyr

Williams, L. E., & Bargh, J. A. (2008). Experiencing Physical Warmth Promotes Interpersonal Warmth. *Science (New York, N.Y.)*, *322*(5901), 606–607. https://doi.org/10.1126/science.1162548

Williams, L. E., Stein, R. , & Galguera, L. (2014). The Distinct Affective Consequences of Psychological Distance and Construal Level. *Journal of Consumer Research, 40*(6), 1123–1138. https://doi.org/10.1086/674212

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods, 51*(1), 409–428. https://doi.org/110.3758/s13428-018-1055-2

Correspondence author: Josef Kundrát, Katedra psychologie, Filozofická fakulta, Ostravská univerzita, Reální 1476/5, 701 03 Ostrava, Czech republic.
E-mail: josef.kundrat@osu.cz

Kundrát, J., Rečka, K., Paulík, K., Baumgartner, F., Malůš, M., Benešová, T., & Thielová, K. (2022). Assessing the attitudes of students toward school subjects with the semantic differential and interactive visual metaphors. *Psychologie a její kontexty, 13*(2), 2022, 63–79. https://doi.org/10.15452/PsyX.2022.13.0011