International Journal of Evaluation and Research in Education (IJERE)
Vol. 10, No. 1, March 2021, pp. 237~244
ISSN: 2252-8822, DOI: 10.11591/ijere.v10i1.20716

Effectiveness of anchoring vignettes in re-evaluating self-rated social and emotional skills in mathematics

Khajidmaa Otgonbaatar
Graduate School of Humanities and Social Sciences, Hiroshima University, Japan
Mongolian Institute for Educational Research, Mongolia

Article Info

Article history:

Received May 4, 2020
Revised Dec 19, 2020
Accepted Jan 27, 2021

Keywords:

Anchoring vignettes
Cronbach's Alpha
Response bias
Social and emotional skills

ABSTRACT

In current assessment practice, self-ratings and questionnaires are a dominant tool used to measure the skills called social and emotional skills or non-cognitive skills, although the tools are affected by various biases. In this regard, the anchoring vignette approach was introduced against the biases, correcting individuals' self-rated responses based on their rating of hypothetical individuals in the scenario. Drawing from students' self-rated social and emotional skills in mathematics, this paper presented the study which examined the effect of anchoring vignette approach on reliability and correlation by comparing self-rated and vignette-corrected scales. Research participants were Mongolian students in ninth grade (N=308). The participants were administered in two scales: self-ratings for math perseverance and cooperative learning in math, followed by a vignette set. The vignette-corrected scale showed higher reliability than the self-rating scale for both math perseverance and cooperative learning in math. Besides, the vignette-corrected scales for gender and region showed a stronger and more significant correlation than the self-rating scales, suggesting that there might be gender- and region-related differences in the way the students act in math class. In summary, these findings suggest that the anchoring vignette approach has the potential to measure social and emotional or non-cognitive aspects of mathematics in a more reliable way. Future studies could further investigate the reliability and validity of the anchoring vignette approach by including more cultural groups and designing the vignettes while considering various math contents and vignette gender.

Corresponding Author:

Khajidmaa Otgonbaatar
Graduate School of Humanities and Social Sciences
Hiroshima University
1-1-1 Kagamiyama, Higashi-Hiroshima City, Hiroshima, 739-8524 Japan
Email: khajidmaaotgonbaatar@gmail.com

1. INTRODUCTION

In the 21st century with science and technology advancement and social and economic climate, as well as environmental issues and changes in the job market, it requires individuals to have a broader range of skills to live in today's unpredictable society [1]. Therefore, different scholars and organizations have introduced skills under various terms to address the challenges facing today's society. For instance, Trilling and Fadel [2] and Partnership for 21st Century Skills [3] used "21st century skills," placing greater emphasis on innovation, critical thinking, and problem-solving skills. In contrast, Farrington, *et al*. [4], Gutman and Schoon [5] employed "non-cognitive skills" to highlight self-control and perseverance. Many others, such as Durlak, *et al*. [6], the Organization for Economic Co-operation Development (OECD) [1], the Center for

Curriculum Redesign [7] and the Collaborative for Academic, Social, and Emotional Learning [8], preferred "social and emotional skills," which include social skills and self-regulation. Note that these terms have been used interchangeably, with different emphases on certain skills.

The present study focuses on the phrase "social and emotional skills" and uses a definition and framework introduced by OECD [1]. However, the previously mentioned terms will appear interchangeably in the later sections of this study. OECD defined social and emotional skills as:

> *"Individual capacities that are manifested in consistent patterns of thoughts, feelings and behaviors, can be developed through formal and informal learning experiences and influence important socio-economic outcomes throughout the individual's life."* [1, p. 34]

OECD [1] categorized social and emotional skills into three domains: achieving goals (perseverance, self-control, passion for goals), working with others (social communication, team working skills) and managing emotions (self-esteem, optimism, confidence).

To address the challenges identified herein, education plays a crucial role in supporting individuals to develop social and emotional [1]. Therefore, education systems must prioritize and foster such skills by revising national curricula and teaching strategies. Following the demand, OECD countries identified the essential skills for future citizens as well as the need to develop the skills in general education objectives and policy documents [1]. This change has occurred in OECD countries. Many other countries have reported that these skills are integrated across the curriculum [9]. Indeed, it is impossible to develop social and emotional skills through a single discipline. Therefore, it is necessary to embed the skills across the disciplines, which leads to revisiting the initial concept of the subjects. For these reasons, in the 21st century—like never before—mathematics education should not be limited to only fundamental mathematics skills but should also include social and emotional skills to contribute to preparing future citizens capable of meeting the identified challenges better.

Aligning with the global trend, the Mongolian government launched reforms in the education sector in 2013. It introduced a new curriculum at the primary level in 2014, a lower secondary level in 2015, and an upper secondary level in 2016. The concept notes on the lower secondary education curriculum states that:

> *"Nowadays, many countries have been developing educational policies aiming to prepare citizens who are capable of being flexible and adaptable to science and technology advancement and to live in an open society in the future. Therefore, the lower secondary curriculum will be designed within the goal of developing "patriotic" Mongolian citizens who are creative, confident, and proficient in decision-making, cooperating and lifelong learning."* [10, p. 18]

Concerning social and emotional skills, Mongolia's new mathematics curriculum—specifically, its lower secondary mathematics curriculum—emphasizes effort, aspiration to learn mathematics and cooperative learning, and the use of several criteria to measure them in the assessment objectives [10].

These skills in mathematics are not limited to Mongolia, but rather appear to be a global issue. Notably, the Progress of International Students Achievement (PISA) 2021 Mathematics Strategic Advisory Group highlighted the following:

> *"In recent times, the digitization of many aspects of life, the ubiquity of data for making personal decisions involving health and investments, as well as major societal decisions to address areas such as climate change, taxation, governmental debt, population growth, the spread of pandemic diseases and the global economy, have reshaped what it means to be mathematically competent and prepared to be a thoughtful, engaged, and reflective citizen."* [11, p. 3]

Thus, it is necessary to reconsider mathematical competencies that initially covered fundamental arithmetic operations, such as addition, subtraction, multiplication, and the division of whole numbers, decimals, and fractions [11]. Therefore, PISA 2021 agreed to expand the model of mathematical literacy by adding two sub-scores, communication (6) and persistence (4), as 21st-century factors [11]. Based on the discussion thus far, perseverance and cooperative learning are two important social and emotional skills in mathematics, which will be explored in the current study.

Kyllonen [12] stated that "… for the past ten years or so, there has been a growing appreciation for the importance of skills other than the cognitive skills," at the same time educational measurement field has

been challenged to measure the skills in a reliable way. Similarly, it was noted that there had been an urgent need to assess non-cognitive skills in the education field [9].

In current measurement practice, most studies that attempted to measure social and emotional skills have relied heavily on questionnaires and self-ratings. However, self-ratings and questionnaires are threatened by "response biases" [13]. Paulhus [13] noted that "… a respondent might choose the option that is most extreme or most socially desirable." For instance, when students respond to a self-rating item like "I am a hard worker" they may opt for higher ratings to hide their weaknesses or attract teachers [14]. Therefore, this social desirability bias affects not only the true level of an individual response but also the whole distribution [14]. More importantly, this bias has the potential to cause a negative influence on the reliability and validity of scores [15].

To address the challenges identified thus far, King and associates [16] developed the anchoring vignettes approach in 2004. In general, anchoring vignettes (AV) are short descriptions about imaginary individuals that illustrate different levels of skills or trait [17, 18]. Respondents are asked to evaluate the fictitious individuals described in the vignettes using the same scale they used to evaluate themselves. The descriptions of the individuals in the vignettes are associated with the skills or characters being evaluated by self-rating items [19] so that respondents' self-ratings can be adjusted based upon their vignettes' evaluation.

King and associates originally developed this approach to measure political attitude [16, 20], but it has since been widely applied in other research areas, such as health [18, 21-23], personality [15-17], customer satisfaction [24], and job satisfaction [25]. Vonkova and Hrabak [19] pointed out that the number of research contributions on AV in the educational field is very limited, although this approach has been applied in many research areas. In particular, this approach has been used to measure parents' satisfaction in charter and public schools [26], improve self-reports of ICT knowledge among upper secondary school students [19], analyze self-reports of dishonest behavior among secondary school students [27], and adjust students' response to the degree of their teachers' assistance [28].

Despite the considerable number of studies in social sciences that provided evidence to encourage the use of the AV, the approach is relatively new in the educational field; it is especially novel in mathematics education with the exception of a PISA 2012 study [29]. Therefore, more studies, focusing on its reliability and validity, need to be carried out in education research. In this regard, the present study aims to examine the effect of the AV approach on reliability and correlation by comparing self-rated, and the AV corrected scales for math perseverance and cooperative learning in math.

2. RESEARCH METHOD

Data were collected through an ongoing research purposing to measure social and emotional skills in mathematics education in Mongolia. Altogether, 308 ninth-grade students from eight schools located in urban and rural areas participated in this study. Table 1 summarizes the socio-demographic characteristics of the sample. The data were gathered during regular class time (45 minutes) in September and October of the 2019–2020 academic year. The data was collected using a paper-and-pencil version of the items and entered the responses into Microsoft Excel for analysis using SPSS 20.0 and R studio 1.2.1335.

Table 1. Social demographic characteristics of the study sample

Characteristics	%
Individual characteristics	
Male	49.0
Female	51.0
Ethnicity	
Mongolian	88.0
Kazakh	12.0
Regional characteristics	
Urban	37.7
Rural	62.3

2.1. Measures

The present study used three self-rating items for each of two scales: math perseverance and cooperative learning in math. These were followed by two vignette sets. The self-rating items for cooperative learning in math were adopted from PISA 2003 [30] (e.g., "I think that mathematics is about working together with others to solve problems") and the items for math perseverance were adopted from Grootenboer and Marshman [31] (e.g., "I usually keep trying with a difficult problem until I have solved it") and Kusmaryono, *et al*. [32] (e.g., "I feel challenged to work hard to find a solution when I get to have a difficult mathematics problem"). The participants were asked to state their agreement to self-ratings using a

5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The participants were then asked to rate imaginary students in the vignette sets using the same scale as with the self-ratings. The vignette sets were developed corresponding to self-rating items. First, the vignette sets were written in English and then translated into Mongolian. The translated versions of the vignette sets were checked by a language expert bilingual in Mongolian and English to ensure linguistic and cultural validation. The vignette sets were written and described as shown in Table 2.

Table 2. Vignette sets for cooperative learning in math and math perseverance

Scale	Vignettes
Cooperative learning in math	Vignette 1 (Low): Bataa tends to argue with others and often start squabbles. So, he prefers to work on math on his own, even if he is stuck with a math problem. He thinks that working with others in math class doesn't help to perform better in mathematics. How much do you agree with the statement "Bataa is a cooperative learner in solving mathematics problems"?
	Vignette 2 (Medium): Solongo doesn't really like working in a group during math class, but sometimes she thinks it is helpful to discuss ideas with others when she is stuck on a math problem. How much do you agree with the statement "Solongo is a cooperative learner in solving mathematics problems"?
	Vignette 3 (High): Jargal finds it easy to cooperate with others to solve a math problem. He enjoys helping others to work well in a group during math class and learning how others solve math problems. As a result, Jargal thinks he does better in mathematics when he works with other students. How much do you agree with the statement "Jargal is a cooperative learner in solving mathematics problems"?
Math perseverance	Vignette 1 (Low): Tuya easily feels desperate and gives up quickly if she encounters difficulty when solving mathematics problems. She is unaware of resources to help her to solve the math problem. How much do you agree with the statement "Tuya is persistent in solving mathematics problems"?
	Vignette 2 (Medium): Tulgaa tries to solve math problems when the answers or solutions are not easily obtainable but gives up when the problem is too complicated. He doesn't put enough effort into solving math problems. How much do you agree with the statement "Tulgaa is persistent in solving mathematics problems"?
	Vignette 3 (High): Oyunaa stays on a math task no matter how difficult it is to find the answers. She searches for more additional information to clarify the problem when she encounters difficulty solving a math problem. Oyunaa always keeps trying with a difficult math problem until she has solved it. How much do you agree with the statement "Oyunaa is persistent in solving mathematics problems"?

2.2. Correcting self-ratings using anchoring vignettes

The present study used a simple non-parametric approach to correct a respondent's self-rating based on his/her vignette response [20]. As previously described, we used three vignettes (low, medium, and high) for each scale; however, any number of vignettes can be used according to the rule $2J + 1$, where J is the number of vignettes [20]. Let the vignettes be Z_1(low), Z_2(medium), and Z_3(high) and the self-rating items be Y_1, Y_2, and Y_3, respectively. The respondents are asked to rate themselves and the vignettes on a 5-point Likert-type scale. Then ith respondents' self-rating, $Y_i \in \{1,2,3,4,5\}$, is rescaled into a vignette-corrected new score, $C_i \in \{1,2,3,4,5,6,7\}$, by applying the following rules [20, 33]:

$C_i = 1$ if $Y_i < Z_1$
$C_i = 2$ if $Y_i = Z_1$
$C_i = 3$ if $Z_1 < Y_i < Z_2$
$C_i = 4$ if $Y_i = Z_2$
$C_i = 5$ if $Z_2 < Y_i < Z_3$
$C_i = 6$ if $Y_i = Z_3$
$C_i = 7$ if $Y_i > Z_3$

Note that the rule allows us to obtain a scalar value of C only when the vignettes are correctly ordered ($Z_1<Z_2<Z_3$) [20]. However, in some cases, responses to the vignettes do not follow the intended order of vignettes, such as when responses to the vignettes are tied ($Z_1<Z_2=Z_3$) or incorrectly ordered ($Z_2<Z_1<Z_3$) [20, 34]. In these cases, King and Wand [20] suggested an interval value of C instead of a scalar value.

To deal with this issue, we used a treatment suggested by Kyllonen and Bertling [35]. If there are ties in the vignette responses, the lowest value among the possible range of values should be chosen. For instance, if $Y_i = Z_1<Z_2=Z_3$ and the possible C value ranges from {4,5,6} between medium and high vignettes, then the lowest value should be selected, which is 4. If there are incorrectly ordered rankings in the responses to the vignettes, the incorrect orderings should be reorganized into complete ties of all vignettes [35]. The tie

should be made at the value the respondent provided to the high vignette, and the same treatment should then be applied to analyze the created tie, as previously explained [35].

3. RESULTS AND DISCUSSION

First, the vignettes' orderings for both scales examined using the "anchor" package for the R program, version 3.0-8 [36]. In Table 3, the first row presents "1,2,3" as the most common ordering as 185 respondents (60%) for cooperative learning in math, and 188 respondents (61%) for math perseverance rated the vignettes as intended. In the second row, for cooperative learning in math, "1,{2,3}" was the second most common ordering, with 59 respondents (19%) tying vignettes 2 and 3. Violations in vignette orderings are presented in rows 4, 5, 6, 8, 9, and 10 for cooperative learning in math and rows 4, 5, 7, 8, 9, and 10 for math perseverance. However, the violation in the ordering appeared in less than 10% of the sample (8.4% for cooperative learning in math and 8.1% for math perseverance), suggesting no problematic vignettes; thus, "it can be treated as measurement error" [17].

Table 3. Vignette orderings (N=308)

Cooperative learning in math				Math perseverance			
Order	Frequency	%	N violation	Order	Frequency	%	N violation
1,2,3	185	0.600	0	1,2,3	188	0.610	0
1,{2,3}	59	0.191	0	{1,2},3	59	0.191	0
{1,2},3	35	0.113	0	1,{2,3}	31	0.100	0
1,3,2	7	0.022	1	2,{1,3}	8	0.025	1
2,1,3	7	0.022	1	2,1,3	8	0.025	1
2,{1,3}	6	0.019	1	{1,2,3}	4	0.012	0
{1,2,3}	3	0.009	0	1,3,2	4	0.012	1
3,1,2	3	0.009	2	{1,3},2	2	0.006	1
{2,3},1	2	0.006	2	3,{1,2}	2	0.006	2
{1,3},2	1	0.003	1	{2,3},1	1	0.003	2

Author examined the vignette equivalence assumption, which posits that "all respondents understand the level of the variable represented in the vignette in the same way" [20]. Table 4 indicates the means and standard deviations for self-rating and each of the three vignettes. The means of the vignettes show consistency between the vignette orderings and characterization of the vignettes. In other words, on average, the high vignette is rated higher than the medium vignette, which is ranked higher than the low vignette, thereby supporting the assumption of vignette equivalence.

Table 4. Descriptive statistics of self-ratings and vignette ratings

	Cooperative learning in math		Math perseverance		
	M	SD	M	SD	N
Self-Rating	3.87	0.98	3.68	0.99	308
Vignette 3 (High)	4.74	0.73	4.74	0.60	308
Vignette 2 (Medium)	3.16	1.24	2.83	1.18	308
Vignette 1 (Low)	1.34	0.82	1.52	0.95	308

Next, to test the effect of the AV on reliability, we compared Cronbach's α [37] to measure the internal consistency of the original and AV corrected scales. As shown in Table 5, Cronbach's α increased from .69 to .83 for the math perseverance scale and from .65 to .81 for the cooperative learning in math scale after the vignette correction. Therefore, the vignette corrected scales showed higher reliability than the original scales for both math perseverance and cooperative learning in math. This result concurs with the finding of a personality study carried out by Primi, *et al.* [15], which discovered that internal consistency increased after the AV correction. The evidence suggested that the AV approach has the potential to improve the reliability of self-rated social and emotional skills in mathematics.

Table 5. Cronbach's α for the original and AV corrected scale

	Cronbach's α for each scale	
	Cooperative learning in math	Math perseverance
Original scale (self-rating)	.654	.695
Corrected scale (AV corrected)	.816	.832

Author then examined whether the AV approach changes the correlation between the scales and the demographic variables, but failed to find a significant relationship between the original scale and other demographic variables for either cooperative learning in math or math perseverance. For the AV corrected scale, it was found a significant correlation between cooperative learning in math and gender as shown in Table 6. This finding suggests there might be gender-related differences in students' learning strategies in math class; wherein females are more cooperative than males.

Table 6. Correlation between cooperative learning in math and demographic variables

N=308	Original scale	AV corrected scale	Gender	Ethnicity	Region
Original scale	1				
AV corrected scale	.667**	1			
Gender	-.077	-.139*	1		
Ethnicity	.047	-.043	-.017	1	
Region	-.059	.027	-.082	-.081	1

Note. *p<0.05; **p<0.01

Again, author found a significant relationship between the AV corrected math perseverance scale and gender as well as region as shown in Table 7. This evidence indicates that there might be gender- and region-related differences in math perseverance among Mongolian ninth-grade students. In other words, females are more persistent than males when solving math problems, and students from rural schools are more persistent than their peers from urban schools.

Table 7. Correlation between math perseverance and demographic variables

N = 308	Original scale	AV corrected scale	Gender	Ethnicity	Region
Original scale	1				
AV corrected scale	.660**	1			
Gender	-.075	-.116*	1		
Ethnicity	.050	.001	-.017	1	
Region	.073	.113*	-.082	-.081	1

Note. *p<0.05; **p<0.01

Based on these findings, it can be inferred that females tend to underreport their level of persistence in solving math problems and learning strategy in mathematics, while students from rural schools are likely to understate their level of persistence in solving math problems when they respond to self-rating items.

Despite the findings, as previously discussed, this study faced some limitations that are worth nothing. First, we emphasized only psychometric properties and the statistical effect of the AV approach rather than its effect on cross-cultural comparisons. However, the original idea of the approach was to compare survey responses across different groups. A cross-cultural comparison would require a larger sample from more cultural groups. This study included only two cultural groups: Kazakhs and Mongolians. Kazakh is a minority ethnic group in Mongolia. Kazakh students speak their mother language at home; however, the main language of instruction at school is Mongolian. Second, the vignette sets were administered only in Mongolian, which means Kazakh students were given Mongolian language vignette sets. The language of the vignettes has some potential to affect the response style. Interestingly, most of the order violations came from the Mongolian group. On the contrary, only one order violation for the vignette set of cooperative learning in math and three order violations for the vignette set of math perseverance were reported by the Kazakh group. Generally speaking, the vignettes functioned well, reporting order violations in fewer than 10% of both scales, indicating that the majority of the respondents understood the vignette scenario in the same way. However, it is challenging to design high-quality vignettes [16]. Third, researchers have stressed that the non-parametric approach faces some disadvantages in dealing with order violations [19]. To address this limitation of the non-parametric approach, we applied the treatment suggested by Kyllonen and Bertling [35] and analyzed order violations as ties. However, this led to a waste of information [19]. Finally, we did not consider the gender of the imaginary individuals in the vignette scenarios. Some studies have suggested that vignette evaluation can be affected by the gender of the fictitious individuals in the vignettes [38].

4. CONCLUSION

This study examined the effect of the non-parametric AV approach on reliability and the correlation between correcting self-rated cooperative learning in math and math perseverance scales. In summary, we found that AV-corrected scales exhibited greater reliability than the original scales, which suggested that the AV approach has the potential to improve the reliability of self-rated social and emotional skills in mathematics. In addition, the AV-corrected scale for gender showed a significantly stronger correlation than the original scales for both cooperative learning in math and math perseverance, suggesting that there might be gender-related differences in the way the students act in math class. Furthermore, the study found a significant relationship between the AV-corrected math perseverance scale and region, whereas the original scale failed to find a significant correlation between the variables. Finally, the present study provided new insights into measures of social and emotional or non-cognitive aspects in mathematics education.

In light of the limitations and considerations, future studies should use the parametric solution of the AV approach, called the compound hierarchical ordered probit (CHOPIT) model, to overcome the limitation of the non-parametric approach. Future studies should also include more cultural groups and socio-demographic variables and design the vignettes while considering various math contents, language, and vignette gender to contribute further theoretical and pedagogical considerations that foster and measure social and emotional skills in mathematics in a more reliable and valid way.

ACKNOWLEDGEMENTS

The author thanks the Japan International Cooperation Agency (JICA) for funding this research and Kh. Amarjargal, English teacher, "New-Era" International School with Cambridge Standards for validating translation of the tools.

REFERENCES

[1] The Organization for Economic Co-operation and Development, "Skills for Social Progress: The Power of Social and Emotional Skills, OECD Skills Studies," 2015. [Online]. Available: https://doi.org/10.1787/9789264226159-en. (accessed Apr. 25, 2020).
[2] B. Trilling and C. Fadel, *21st century skills*. San Francisco, Calif: Jossey-Bass, 2013.
[3] Partnership for 21st Century Skills, "P21 Framework Definitions," 2009. [Online]. Available: https://eric.ed.gov/?id=ED519462. (accessed Apr. 17, 2020).
[4] Farrington, C.A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, TS., Johnson, D.W., and Beechum, N.O., "Teaching Adolescents To Become Learners The Role of Noncognitive Factors in Shaping School Performance: A Critical Literature Review," 2012. [Online]. Available: https://www.academia.edu (accessed Apr. 15, 2020).
[5] L. Gutman and I. Schoon, "The impact of non-cognitive skills on outcomes for young people," 2013. [Online]. Available: https://pdfs.semanticscholar.org. (accessed Apr. 15, 2020).
[6] J. Durlak, R. Weissberg, A. Dymnicki, R. Taylor and K. Schellinger, "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions," *Child Development*, vol. 82, no. 1, pp. 405-432, 2011, doi: 10.1111/j.1467-8624.2010.01564.x.
[7] Center for Curriculum Redesign, "Framework of Competencies/Subcompetencies," 2020. [Online]. Available: https://curriculumredesign.org/framework/ (accessed Apr. 15, 2020).
[8] Collaborative for Academic, Social, and Emotional Learning, "Core SEL Competencies," 2017. [Online]. Available: https://casel.org/core-competencies/ (accessed Apr. 15, 2020).
[9] The Ontario Public Service, "21st Century Competencies," 2016. [Online]. Available: http://www.edugains.ca. [Accessed: 17-Apr-2020].
[10] Ministry of Education, Culture, Science and Sports - Mongolia, "Lower Secondary Core Curriculum," 2015. [Online]. Available: https://mecss.gov.mn/news/94/ (accessed Apr. 12, 2020).
[11] The Organization for Economic Co-operation and Development, "PISA 2021 Mathematics: A Broadened Perspective," 2017. [Online]. Available: https://www.oecd.org/ (accessed Apr. 18, 2020).
[12] P. C. Kyllonen, "Measurement of 21st Century Skills Within the Common Core State Standards," Presented at Invitational Research Symposium on Technology Enhanced Assessments, 2012, May 7-8, Princeton, USA, 2012.
[13] D. Paulhus, "Measurement and control of response bias," *Measures of Personality and Social Psychological Attitudes*, pp. 17-59, 1991.
[14] M. West, *et al*., "Promise and Paradox: Measuring Students' Non-Cognitive Skills and the Impact of Schooling," *Educational Evaluation and Policy Analysis*, vol. 38, no. 1, pp. 148-170, 2016, doi: 10.3102/0162373715597298.
[15] R. Primi, C. Zanon, D. Santos, F. De Fruyt, and O. John, "Anchoring Vignettes," *European Journal of Psychological Assessment*, vol. 32, no. 1, pp. 39-51, 2016, doi: 10.1027/1015-5759/a000336.
[16] G. King, C. Murray, J. Salomon, and A. Tandon, "Enhancing the validity and cross-cultural comparability of measurement in survey research," *American Political Science Review*, vol. 98, no. 1, pp. 191-207, 2004, doi: 10.1017/s000305540400108x.
[17] S. Weiss and R. Roberts, "Using Anchoring Vignettes to Adjust Self-Reported Personality: A Comparison Between Countries," *Frontiers in Psychology*, vol. 9, 2018, doi: 10.3389/fpsyg.2018.00325.

[18] J. Pacheco, "Using Anchoring Vignettes to Reevaluate the Link between Self-Rated Health Status and Political Behavior," *Journal of Health Politics, Policy and Law*, vol. 44, no. 3, pp. 533-558, 2019, doi: 10.1215/03616878-7367060.
[19] H. Vonkova and J. Hrabak, "The (in) comparability of ICT knowledge and skill self-assessments among upper secondary school students: The use of the anchoring vignette method," *Computers & Education*, vol. 85, pp. 191-202, 2015, doi: 10.1016/j.compedu.2015.03.003.
[20] G. King and J. Wand, "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes," *Political Analysis*, vol. 15, no. 1, pp. 46-66, 2007, doi: 10.1093/pan/mpl011.
[21] H. Grol-Prokopczyk, "Age and Sex Effects in Anchoring Vignette Studies: Methodological and Empirical Contributions," *Survey Research Methods*, vol. 8, no. 1, pp. 1-17, 2014. [Online]. Available: https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC4302337&blobtype=pdf (accessed Apr. 18, 2020).
[22] A. Hinz, W. Häuser, H. Glaesmer and E. Brähler, "The relationship between perceived own health state and health assessments of anchoring vignettes," *International Journal of Clinical and Health Psychology*, vol. 16, no. 2, pp. 128-136, 2016, doi: 10.1016/j.ijchp.2016.01.001.
[23] B. Poksinska and P. Cronemyr, "Measuring quality in elderly care: possibilities and limitations of the vignette method," *Total Quality Management & Business Excellence*, vol. 28, no. 9-10, pp. 1194-1207, 2017, doi: 10.1080/14783363.2017.1303875.
[24] O. Paccagnella, "Anchoring vignettes with sample selection due to non-response," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 174, no. 3, pp. 665-687, 2011, doi: 10.1111/j.1467-985x.2011.00707.x.
[25] N. Kristensen and E. Johansson, "New evidence on cross-country differences in job satisfaction using anchoring vignettes," *Labour Economics*, vol. 15, no. 1, pp. 96-117, 2008, doi: 10.1016/j.labeco.2006.11.001.
[26] J. Buckley and M. Schneider, *Charter schools: Hope or hype?* Princeton, N.J: Princeton University Press, 2007.
[27] H. Vonkova, S. Bendl, and O. Papajoanu, "How Students Report Dishonest Behavior in School: Self-Assessment and Anchoring Vignettes," *The Journal of Experimental Education*, vol. 85, no. 1, pp. 36-53, 2016, doi: 10.1080/00220973.2015.1094438.
[28] H. Vonkova, G. Zamarro and C. Hitt, "Cross-Country Heterogeneity in Students' Reporting Behavior: The Use of the Anchoring Vignette Method," *Journal of Educational Measurement*, vol. 55, no. 1, pp. 3-31, 2018, doi: 10.1111/jedm.12161.
[29] The Organization for Economic Co-operation and Development, "PISA 2012 technical report," OECD Publishing, Paris: Author, 2014.
[30] The Organization for Economic Co-operation and Development, "PISA 2003 technical report," OECD Publishing, Paris: Author, 2005.
[31] P. Grootenboer and M. Marshman, *Mathematics, Affect and Learning*, 1st ed. Springer Science + Business Media Singapore, 2016.
[32] I. Kusmaryono, S. Hardi and K. Nur, "Evaluation on Dispositional Mental Functions of Cognitive, Affective, and Conative in Mathematical Power Problems-Solving Activity," *Journal of Mathematics Education*, vol. 11, no. 1, pp. 81-102, 2018, doi: 10.26711/007577152790022 (accessed Apr. 17, 2020).
[33] M. von Davier, H. Shin, L. Khorramdel, and L. Stankov, "The Effects of Vignette Scoring on Reliability and Validity of Self-Reports," *Applied Psychological Measurement*, vol. 42, no. 4, pp. 291-306, 2017, doi: 10.1177/0146621617730389.
[34] D. Hopkins and G. King, "Improving Anchoring Vignettes: Designing Surveys to Correct Interpersonal Incomparability," *Public Opinion Quarterly*, vol. 74, no. 2, pp. 201-222, 2010, doi: 10.1093/poq/nfq011.
[35] P. C. Kyllonen and J. P. Bertling, "Draft report: Anchoring vignettes reduce bias in non-cognitive rating scale responses," Princeton, NJ: ETS/OECD, 2014.
[36] J. Wand, G. King and O. Lau, Statistical Analysis of Surveys with Anchoring Vignettes, 2016. Accessed: Apr. 27, 2020. [Online]. Available: https://cran.r-project.org/web/packages/anchors/anchors.pdf
[37] L. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297-334, 1951, doi: 10.1007/bf02310555.
[38] H. Jürges and J. Winter, "Are Anchoring Vignettes Ratings Sensitive to Vignette Age and Sex?" *Health Economics*, vol. 22, no. 1, pp. 1-13, 2011, doi: 10.1002/hec.1806.