

Improving Vignette Descriptions and Question Formats to Measure Distance Vision: Evidence from Cognitive Interviews among Students in China

Yanfang Su¹, Gordon Willis²,
and Joshua A. Salomon³

Abstract

Vignette design has been largely neglected in anchoring vignette studies. This study aimed to contribute to the science of vignette design by developing and evaluating vignettes for measuring vision in rural China. Cognitive interviews were conducted among 36 participants in a Chinese middle school. The respondents either directly evaluated vision of the vignette character (i.e., noncomparative judgment) or compared their own vision with that of the vignette character (i.e., comparative judgment). It was found that a hypothetical person in the vignette was successfully envisioned by

¹ Freeman Spogli Institute for International Studies, Stanford University, Stanford, CA, USA

² National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

³ Harvard T. H. Chan School of Public Health, Boston, MA, USA

Corresponding Author:

Yanfang Su, Freeman Spogli Institute for International Studies, Stanford University, Encina Hall E401, Stanford, CA 94305, USA.

Email: yanfangsu@mail.harvard.edu

participants in grade 7 and beyond. However, more than half the participants were unable to accurately estimate distances expressed in meters. Some participants were critical in self-evaluation, yet tolerant of others' performance. Participants more easily produced an answer and had greater confidence in the answer in comparative judgment than for noncomparative judgment. We conclude with recommendations for designing concise and complete vignettes and suggest the use of comparative rather than non-comparative judgment.

Measuring subjective judgments related to health, through the use of self-report surveys, is challenging, especially because responses across respondents may be difficult to compare. To address this challenge, the anchoring vignette approach has been developed to improve interpersonal comparability of survey responses, especially to questions that use ordered categorical response scales (King et al. 2004; Salomon et al. 2001). The anchoring vignette approach relies on short descriptions of hypothetical persons with different levels on a dimension of interest to gain insights into the ways that respondents reply to particular survey questions related to that dimension. It is hypothesized that by using responses to vignettes to identify systematic differences in the ways that respondents interpret and use the categories on a given response scale, researchers can recalibrate the responses to self-assessment questions (King et al. 2004; Salomon et al. 2001). Anchoring vignettes have been advanced with tailored statistical methods (Wand 2013; Wand et al. 2011) and are used widely in the World Health Survey and in other empirical studies (Chevalier and Fielding 2011; Damacena et al. 2005; King et al. 2012; Salomon et al. 2004; Wada et al. 2011).

However, vignette design, including designing vignette descriptions and question formats, has been largely neglected (Hopkins and King 2010; Kapteyn et al. 2011). More recent studies have revealed substantial violations of measurement assumptions underlying anchoring vignette techniques (Bago d'Uva et al. 2011; Groj-Prokopczyk et al. 2015; Hirve et al. 2013; Kapteyn et al. 2011), and it has been suggested that measurement assumptions are more likely to hold true if the description of the vignette character's condition is complete and concise (Kapteyn et al. 2011). It has also been suggested that there is need for further work on improving vignette descriptions and question formats before vignettes are used in practice (Hopkins and King 2010; Kapteyn et al. 2011).

In the current study, cognitive interviewing was used to evaluate and improve vignette descriptions and question formatting. Cognitive interviewing is an evaluation method used in questionnaire design that allows researchers to identify problems in question formulation that may prevent them from effectively collecting information (Willis 2005). The interviewer applies verbal probing techniques (Forsyth and Lessler 1991) and the participants are allowed to think aloud (Ericsson and Simon 1980) during the interviews. Documenting the cognitive processes in responding to vignette descriptions and question formats is critical for understanding whether the measurement assumptions are satisfied, which is an important requirement for the validity of the anchoring vignette approach.

This study included assessments of two specific vignette descriptions, representing poor or good latent visual acuity. Among different health domains, vision was chosen as the focus of this study because both objective measures and self-assessment can be applicable to measuring distance vision (King et al. 2004), which affords an opportunity to validate self-assessments in relation to a measured standard.

This study also assessed two question formats that involved either non-comparative judgment or comparative judgment. The original form of the anchoring vignette technique asks survey respondents to first self-assess the domain of interest and to then make a separate, noncomparative judgment concerning a hypothetical character presented in a vignette (King et al. 2004; Salomon et al. 2001). The original form is based on two assumptions: vignette equivalence and response consistency (Salomon et al. 2004). Vignette equivalence refers to the requirement that underlying domain levels represented in each vignette are understood in approximately the same way by all. Response consistency refers to the requirement that individuals use similar standards in self-assessment and in the evaluation of vignette scenarios. As an alternative anchoring vignette technique, researchers have proposed a comparative judgment in which the respondent is asked for a direct comparison between the self and the vignette (Hopkins and King 2010). The comparative judgment is also based on the assumption of vignette equivalence but free from the requirement of response consistency. These two competing question formats have been evaluated by other researchers on the topic of rest/energy (Hopkins and King 2010) and are evaluated on the domain of vision in this study.

As one component in a broader measurement study, this study aimed to develop and evaluate vignettes for measuring vision in rural China. This study answers the following research questions: How do participants comprehend, recall, judge, and respond to vignettes representing poor and good

visual acuity? Do participants' comprehension and ability to respond differ between comparative and noncomparative question formats in the vignettes? Were there any differences in responses according to participants' level of visual ability?

Method

Sample

We conducted the study in Jingning Middle School, a public school in a rural area in Zhejiang Province, China. Students in middle schools in China are between grade 7 and grade 12, and the study aimed to enroll participants to represent different educational levels. However, students in grade 12 were preparing for university entrance exams and were not invited to participate in the study. An interviewer selected 36 participants, 18 in grade 7 (with a mean age of 13 years) and 18 in grade 11 (with a mean age of 18 years). Teachers helped identify participants with particular characteristics of interest (e.g., students who had glasses, students who did not need glasses, students who always wore glasses, and students who sometimes wore glasses).

Survey Questions

We tested two initial versions of survey questionnaires in cognitive interviews, and there were three questions in each version (Table 1). Each questionnaire consisted of one self-assessment question, two vignettes, and a preface before each type of question. Two hypothetical persons with common Chinese names, Wang Wu and Zhang San, were introduced in vignettes (Table 1). It was hypothesized that response consistency might be enhanced by specifying information about the vignette characters' age and gender in the preface. Two vignette descriptions represented scenarios in which a hypothetical person had either a distance vision of less than 5 m or greater than 20 m. Only vignette question formats were phrased in different ways within the two versions: comparative and noncomparative judgment (Table 1).

Procedures

This study was reviewed and approved by the institutional review board at Harvard University. Participants were informed about the study goal, procedure, and estimated time. Adult participants' written consent was secured.

Table 1. Two Versions of Survey Questionnaires.

Preface 1	Now, I would like you to think about your own vision without glasses or contact lenses.	
Self-assessment of vision	(1) At the present time, would you say your distance eyesight is 1) Excellent 2) Good 3) Fair 4) Poor 5) Very poor	
Preface 2	When answering the next questions, I want you to think about Wang Wu and Zhang San of your age and gender. Please think about Wang Wu and Zhang San's vision without glasses or contact lenses.	
Vignette questions	<p data-bbox="298 607 610 634">Noncomparative judgment</p> <p data-bbox="298 639 610 769">(2) Wang Wu finds faces to appear blurry at a distance of 5 m. Would you say Wang Wu's distance eyesight is</p> <p data-bbox="298 774 610 911">1) Excellent 2) Good 3) Fair 4) Poor 5) Very poor</p> <p data-bbox="298 915 610 1268">(3) Zhang San can recognize familiar people's faces and pick out facial expression (e.g., angry, smile) at a distance of 20 m quite distinctly. Would you say Zhang San's distance eyesight is</p> <p data-bbox="298 1138 610 1268">1) Excellent 2) Good 3) Fair 4) Poor 5) Very poor</p>	<p data-bbox="625 607 929 634">Comparative judgment</p> <p data-bbox="625 639 929 745">(2) Wang Wu finds faces to appear blurry at a distance of 5 m. Would you say your distance eyesight is</p> <p data-bbox="625 750 929 911">1) Better than Wang Wu's 2) The same as Wang Wu's 3) Worse than Wang Wu's</p> <p data-bbox="625 915 929 1268">(3) Zhang San can recognize familiar people's faces and pick out facial expression (e.g., angry, smile) at a distance of 20 m quite distinctly. Would you say your distance eyesight is</p> <p data-bbox="625 1105 929 1268">1) Better than Zhang San's 2) The same as Zhang San's 3) Worse than Zhang San's</p>

Children under 18 years old were recruited with parental written consent and each child's assent. Interviews were conducted in Chinese by a native speaker, and all interviews were tape-recorded and had notes taken.

Table 2. Study Design.

Step 1	Question appraisal by interviewer
Step 2	First round of study ($n = 18$) (Concurrent cognitive interviewing, followed by objective measure of vision)
Step 3	Refinement of questions
Step 4	Second round of study ($n = 18$) (Retrospective cognitive interviewing, followed by objective measure of distance vision)
Step 5	Finalization of vignettes and question formats

Table 2 presents the design of study procedures. Cognitive pretesting consisted of two rounds. The first round of cognitive interviewing involved concurrent probing (Willis 2005), in which the participant self-administered a survey question, then the interviewer immediately asked additional questions regarding this specific question. The process was repeated for each of the three questions. To provide a more naturalistic environment where the participant's responses to the survey questions were uncontaminated by probing, the second round was retrospective: The participant self-administered the entire written survey that consisted of three questions and submitted it to the interviewer, then the interviewer asked additional questions.

The only exception was that the concept of distance was pretested in a semistructured manner over the course of the study, without two clear-cut rounds. The concept of distance was first introduced as 20 m and 5 m. The interpretation of distance was tested by several probe questions. For example, "What is the length of this interview room?", "What is your best guess of the distance between you and the eye examination chart?", and "How far is 5 m?" After pretesting 20 m among 18 participants, it was revised to 10 m and pretested among 10 more participants, while 5 m remained unchanged in pretesting among 28 participants. After the refinement of vignette descriptions, distance was not expressed in meters for the remaining eight participants.

In pretesting question formats, noncomparative or comparative judgment was randomly presented first to the participant. A further probe on comparative judgment was designed for participants who responded to noncomparative judgment and vice versa. For instance, after conducting an interview involving noncomparative judgment, the interviewer elaborated: "In the previous questions, I asked you to talk about your vision and

then to imagine and talk about the vision of another person, Wang Wu. I would like to know how you would respond if I asked you a more direct comparison.” The participant responded to the direct comparison question, and after that a probe was used: “Which question is easier to answer, in evaluating the vignette character’s vision alone or in direct comparison between self and the vignette character?”

After each interview, the participant received an objective measure of visual acuity by using the “simplified Snellen chart,” consisting of the letter E oriented in different directions. The respondent stood 5 m away from the chart and indicated the direction of the E. The interviewer conducted the test after receiving training from an optometrist in a local setting to ensure safety and accuracy. In the objective measure, distance vision without glasses or contact lenses was measured even if the participant typically wore visual corrections. The values on the simplified Snellen chart ranged from 4.0 to 5.3, with a larger number indicating better vision. The Snellen results were summarized in three categories: worst vision (with Snellen chart value 4.0–4.3), moderate vision (4.4–4.7), and best vision (4.8–5.3).

Data Analysis

We used the four-stage Tourangeau et al. (2000) psychological model of the survey response process as the framework to guide our data analysis. Accordingly, in the Results section, in addition to a description of the participants’ characteristics, we summarized four key challenges in anchoring vignette techniques: comprehension, recall, judgment, and responses to vignette descriptions and question formats.

We synthesized data collected through verbal probing, thinking aloud, and objective measurement of vision. We listened to the taped voice records for direct quotes from different participants, summarized the main challenges in vignette design through qualitative analysis, and used the text summary approach described by Willis (2015) to understand the frequency of each challenge. For example, we explored the systematic differences in the survey response processes among participants as a function of objective vision. Accordingly, we present both the qualitative and quantitative results when applicable.

Findings

Characteristics of the Participants

Table 3 presents the characteristics of the study sample.

Table 3. Summary Statistics.

	Number	%
Gender		
Males	17	47
Females	19	53
Wears glasses		
No	16	44
Sometimes	11	31
Always	9	25
Objective vision		
Worst (4.0–4.3)	11	31
Moderate (4.4–4.7)	10	28
Best (4.8–5.3)	15	42

Note: $N = 36$.

Comprehension

Hypothetical person. Based on content analysis of responses to probes, we determined that all participants were able to successfully envision the hypothetical person presented in the vignettes, with respect to his or her demographic characteristics. Participants commented that Zhang San and Wang Wu were two common names that were used to describe hypothetical characters in the mathematical examinations at the end of semester and represented specific people like the participant himself or herself or a classmate. For instance, one participant reported, “Wang is the same as me, except eyesight.”

Challenges to vignette equivalence. The main challenges in vignette equivalence were the understanding of distance and descriptions in the vignettes (i.e., appear blurry, familiar people’s face, angry). We discuss these issues in depth below.

The majority of participants experienced difficulty in estimating these distances, especially for the 20 m description. Among participants who were asked to estimate the distance of 5 m through cognitive probing ($n = 28$), 14 of the 28 participants (50%) accurately estimated that distance, and the estimation task became more challenging as the distance increased, with accurate estimates received from only four of the 10 participants (40%) responding about 10 m and four of the 16 participants (25%) responding about 20 m. (As the interviews were semistructured, for two out of the 18 students, the interviewer didn’t probe about their estimations of 20 m.) Because of the difference in estimated distance, vignettes described with

numeric distance (e.g., 20 m) were unlikely to be interpreted equivalently among participants.

Participants also reported varied comprehension of the described vignettes (i.e., [Wang Wu] finds faces to appear blurry at a distance of 5 m, [Zhang San] can recognize familiar people's faces and pick out facial expression [e.g., angry, smile] at a distance of 20 m quite distinctly). Overall, six out of the 18 participants (33%) exhibited difficulty with the phrase "appear blurry," four out of the 18 participants (22%) reported difficulty with the word "angry," and two out of the 18 participants (11%) reported difficulty with the phrase "familiar people's face." For example, appear blurry was understood in different ways: (1) Wang cannot recognize the person, (2) Wang can recognize the person but Wang experienced a double image of the face, and (3) Wang cannot pick up the details such as moles on the face. Furthermore, some participants wanted to know more about the details regarding the blurriness.

Participants commented on "familiar people's face" in a richly descriptive manner. One participant commented that it was easier to recognize a familiar person's face than a stranger's; accordingly, he was struggling to rate Zhang's vision (i.e., representing good vision) as "good" or "excellent."

Recall and Judgment

Recall. Participants with poor vision had a difficult time recalling a sense of good vision. In evaluating the good-vision vignette in comparative judgment, one participant commented, "For the vignette with distance of 10 meters, I can hardly recall my own experience because I became near-sighted since grade 4 and I have no memory of visual clearness at 10 meters."

The importance of a reference point in making a judgment. Interestingly, some participants utilized a previous vignette to make a judgment regarding the next vignette. For instance, 20 m was assessed with reference to an understanding of 5 m. One participant said, "If the vignette with 20 meters is presented first, I am not sure how to answer it; maybe Zhang's vision is excellent or good. Even though I don't have the absolute sense of 20 meters, I know how far it should be, compared with 5 meters."

As a key feature in comparative judgment, participants used the self as the reference point to make relative judgments. Because most participants had a good sense of their own vision, it was easier to make a judgment by

comparing themselves and the vignette's character. Further, in comparative judgment, relevance to one's own life experience appeared to make the judgment process more engaging. However, for noncomparative judgments, participants needed to generate an absolute sense about distance or seek a new reference point for further judgment. Given that relevance to the participants' daily life was an important factor to consider in designing vignettes, one challenge in comparative judgment was the limits of experience. Participants with good vision had limited experiences like that of Wang's poor vision.

Boundary defined in vignette descriptions. Several participants recommended defining the vignette character's visual capacity in terms of what can and cannot be seen clearly. One participant stated, "I am not sure about Wang's vision if the only thing I know is that he is my age and has difficulty seeing clearly at a distance of 5 meters. Does he see clearly at the distance of 2 meters?" The participant found it helpful to add, "But Wang can see clearly at the distance of 2 meters."

Response Processes

Challenges to response consistency. In the first round of interviews, eight (44%) of the 18 participants reported that the information on vignette characters' gender was irrelevant, and 10 participants (56%) thought the opposite. The information about gender was tentatively removed after the first round of interviews, and in the second round of interviews, among 18 participants, none asked whether the hypothetical person was a girl or a boy participant.

We found that the assumption of response consistency (i.e., that the same standards are applied to the self as to the vignette character) may have been violated by a tendency within Asian culture to be critical in self-evaluation but tolerant of others' performance. Four of the 36 participants (11%) had inconsistent standards in rating their own vision and the vignette's vision. The following is an example in which a participant rated his vision as good in self-assessment:

Interviewer: "How sure are you that your vision is good?"

Participant: "In most of the cases, I would say it is good. I think my vision is good among the classmates. If you ask me to compare with another classmate, I might be the better one. But, I like to be modest."

Table 4. Difficulty in Responding to Vignettes.

Response Difficulty	<i>n</i>	%
No difference	9	29
Poor vision vignette is more difficult	12	39
Good vision vignette is more difficult	10	32

Note: $N = 31$.

The participant rated Wang's vision as "poor." He thought the question was easy to answer:

Participant: "Wang's vision might be very poor . . . or poor."

Interviewer: "Really? Why did you rate his vision as 'poor'?"

Participant: "Probably, . . . I should not rate others' vision as 'very poor'."

After receiving the highest possible score on the Snellen chart, the participant insisted on the original answer and he was sure he would choose good as the answer. When the participant was asked about the reason, he said, "If others' [vision] is very poor, I'd like to be lenient. I seldom rated others' performance as 'very poor.' For myself, I would like to be humble."

Responding to different vignettes. Thirty-one out of the 36 participants answered the question regarding difficulty in responding to vignettes. Among those 31 participants, nine (29%) thought the two vignettes were equivalent in response difficulty, 12 (39%) thought that it was more difficult to answer the poor-vision vignette, and the other 10 (32%) thought the opposite (Table 4).

After categorizing participants who differentiated the vignettes into three groups by the level of objective vision, a clear pattern emerged (Table 5). All participants with worst objective vision ($n = 8$) and four out of the six participants (67%) with moderate objective vision found the poor-vision vignette more difficult. Meanwhile, all participants with best objective vision ($n = 8$) found the good-vision vignette more difficult (Table 5).

Further, some participants found the use of the second vignette redundant: "I can only see clearly at a distance of 2 meters. My vision is worse than Wang's and definitely worse than Zhang's. After answering the question about Wang's vision, I found the sequential question about Zhang's vision not interesting at all."

Table 5. Differentiated Response Difficulty by Objective Vision.

Problems	N	Poor-Vision Vignette (Wang)		Good-Vision Vignette (Zhang)	
		n	%	n	%
More difficult to answer the question	22	12	55	10	45
Worst objective vision	8	8	100	0	0
Moderate objective vision	6	4	67	2	33
Best objective vision	8	0	0	8	100

Note. $N = 22$.

Table 6. Difficulty in Responding to Question Formats.

Response Difficulty to Question Formats	n	%
No difference	6	17
Noncomparative judgment is more difficult	23	66
Comparative judgment is more difficult	6	17
Total	35	100

Note: $N = 35$.

Responding to different question formats. Among 35 responding participants, most (66%) reported that noncomparative judgment was more difficult (Table 6). The remaining participants were equally divided between reporting that comparative judgment was more difficult (17%) and no difference between the two approaches (17%).

The statement that noncomparative judgment was more difficult was consistent among different objective vision groups (Table 7). Among 32 responding participants, more were uncertain about their answer when using noncomparative judgment.

Difficulty in estimating distance was more evident for noncomparative judgments than for comparative judgments. For instance, in comparative judgment, one participant stated, “I cannot see clearly at a distance of 1 meter and it is inferred that Wang’s vision is better than mine, even though I don’t know how far 5 meters is.” For the same participant, the noncomparative judgment became challenging because of the uncertainty about how far 5 m was.

Among participants who reported noncomparative judgment to be easier than comparative judgment, one participant’s comment was representative:

It is very straightforward to rate Wang's vision. When I am asked about the comparison between Wang's vision and mine, I first think about Wang's vision, and then I evaluate my own vision. I have to think back and forth when doing the comparison, but very quickly came to a conclusion when rating Wang's vision.

Table 7. Problems in Responding to Question Formats.

Problems	N	Comparative Judgment		Noncomparative Judgment	
		N	%	n	%
1. More difficult to answer the question	29	6	21	23	79
Worst objective vision	11	3	27	8	73
Moderate objective vision	8	1	13	7	88
Best objective vision	10	2	20	8	80
2. Uncertainty of each answer	32	4	13	15	47

Finalized Vignettes and Question Format

The recommended vignettes were:

In the cafeteria, Xiao Wang can clearly recognize students sitting at his table, but not those sitting at the next table.

From the last row in the classroom, Xiao Zhang can clearly recognize his teacher, but not the small written text on the blackboard.

The finalized question format was:

Would you say your distance eyesight is better than, the same as, or worse than Xiao Wang's?

Would you say Xiao Wang's distance eyesight is excellent, good, fair, poor, or very poor?

Discussion and Limitations

This study documented the challenges in comprehension, recall, judgment, and responses to vignette descriptions and question formats among participants with different objective levels of visual ability in rural China using cognitive interviewing. This study contributes to the literature relating to

both the use of cognitive interviewing as a pretesting and evaluation procedure, particularly in the development and evaluation of vignettes.

Designing concise, yet complete vignette descriptions is clearly challenging (Kapteyn et al. 2011). Overall, participants were able to successfully envision the hypothetical person with respect to his or her demographic characteristics. In the domain of distance vision, it was suggested that participants' evaluation of vignettes was affected by knowledge of vignette characters' age but not gender. Given that the use of Chinese common names is not gender-specific (e.g., Zhang San and Wang Wu), omitting hypothetical characters' gender might be an advantage in applying vignette methods in a Chinese context. It was suggested by Grol-Prokopczyk and colleagues (2011) that omission of information about a vignette character's gender is not feasible in most linguistic settings. On the other hand, one study in the United States and the Netherlands suggested that participants responded differently to vignettes with a female name than with a male name (Kapteyn et al. 2007). Therefore, vignette equivalence may not hold, at least if the potentially subtle connotations of vignette persons' names are not fully controlled (Jürges and Winter 2013). There is probably no need to include male or female pronouns or first names with implied gender when using anchoring vignette techniques in Chinese, at least where the topic is visual acuity.

Ambiguous phrases and lengthy wording caused difficulty in comprehension. More than half of the participants were unable to accurately estimate distances expressed in 5, 10, and 20 m. The ability to accurately assess distance varied substantially among participants, which reduced vignette equivalence. In a study measuring mobility level in Asian countries, interpretations of mobility level, as presented in the vignette, varied significantly among participants across countries (Hirve et al. 2013). Distance expressed in meters, used in the World Health Survey vignettes on vision, seemed to cause difficulties in comprehension, which contrasts with a conventional assumption that "objective" details should reduce cross-cultural misunderstanding (Pasick et al. 2001). It was found from responses to different vignettes that, while the vignette described a scenario that was similar to the respondent's, it became more difficult to answer, compared with a case that diverged more from the respondent's own experience.

This study recommended the use of comparative judgment. Hopkins and King (2010) recommended noncomparative judgment because they found significant ranking inconsistency and center-seeking tendency in comparative judgment in measuring political efficacy and rest/energy. However, Su (2015) quantitatively tested whether comparative judgment was more valid

than noncomparative judgment in measuring vision and found that ranking inconsistency was nonsignificantly different between comparative judgment and noncomparative judgment in distance vision.

As a further conclusion of the current study, response inconsistency in noncomparative judgment may occur due to values in Asian culture. However, whereas the current study found that the Asian culture of being “strict with oneself and lenient toward others” existed among respondents, Au and Lorgelly found the opposite result in different populations: Some English-speaking respondents (mostly Australians) were optimistic in self-assessment and rated the vignettes on a different scale (Au and Lorgelly 2014). Response inconsistency involving such differences in cross-cultural patterns would undermine international comparability of vignettes. Therefore, we again suggest using comparative judgment because it does not require response consistency.

Findings from this study add to a broader literature on cultural norms and psychology in an Asian context. Many of the challenges in cross-cultural survey work recognized by researchers (Johnson 1998, 2006; Willis and Miller 2011) can be at least partially attributed to the impact of culture on self-reports because culture exerts a fundamental influence on basic psychological processes (Chiu and Hong 2013; Keesing 1974; Lehman et al. 2004).

It may be effective to ask questions like: “Can you clearly recognize the small written text on the blackboard from the last row in the classroom?” We predict that this version of direct questioning with more specific information might be superior to vignette methods for at least two reasons: (1) there is no involvement of the vignette characters and (2) there is no need to meet the requirement of response consistency. It would be interesting to test this version of the question in future studies.

The study has important limitations. First, the sample size of the cognitive interviews was 36; with a small sample size, the quantitative results were suggestive rather than conclusive. Second, the results might not be generalized to the wider population because a representative sample was not included. Third, it was observed that some Chinese participants were modest in self-assessment of vision in interviews, but it is unknown whether the participants would honestly report or overreport visual acuity in a self-administered anonymous survey in which the interviewer is absent. Fourth, some of the findings are domain-specific or cultural sensitive, and therefore they may not generalize to other domains or cultures, while other findings might be more generalizable. Fifth, this study was conducted in rural China

and the findings might not pertain equally to urban areas, given rural–urban differences in social economic status.

Conclusion and Implications

This study develops vignettes for measuring vision in Chinese middle schools and generated knowledge about the design of vignette descriptions and question formats. In the application of vignette methods, we recommend pretesting vignettes. The study also points to several recommendations for designing concise and complete vignettes for enhancing vignette equivalence. First, we suggest avoiding vague phrases, lengthy wording, and metrics expressed in precise but difficult terms (e.g., distance in meters). Second, vignettes should be described in such a way that the domain of interest is defined both by what *can* be done and what *cannot* be done by the vignette character. Third, there may be a trade-off between collecting data and increasing redundancy by adding more vignettes.

Further, concerning the question format used, we recommend using comparative judgment, as it was easier to reach an answer and to be certain about that answer than for noncomparative judgment. Response inconsistency in noncomparative judgment may occur due to values in Asian culture concerning strictness in self-assessment and tolerance toward others. Comparative judgment eliminates the need to assume response inconsistency, with the respondent's own characteristics providing the reference point in this question format.

Acknowledgments

Our thanks to the Jingning Middle School, China, to the participants who devoted time and effort in the interviews, and to Fangfang Xing, Huangzhong Ji, and Jiafeng Wu for their logistical support. Thanks to William Hsiao, Margaret McConnell, Gary King, Ben Campbell, Yang Cai, Xiaoyu Pan, and Anne Watt for their valuable comments. We are indebted to the B&P Foundation for their financial support.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors received financial support

from the B&P Foundation to conduct research. The authors received no financial support for authorship or publication of this article.

References

- Au, N., and P. K. Lorgelly. 2014. Anchoring vignettes for health comparisons: An analysis of response consistency. *Quality of Life Research* 23:1721–31.
- Bago d’Uva, T., M. Lindeboom, O. O’Donnell, and E. Van Doorslaer. 2011. Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity. *Journal of Human Resources* 46:875–906.
- Chevalier, A., and A. Fielding. 2011. An introduction to anchoring vignettes. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174:569–74.
- Chiu, C.-Y., and Y.-Y. Hong. 2013. *Social psychology of culture*. New York: Psychology Press.
- Damacena, G. N., M. T. L. Vasconcellos, and C. L. Szwarcwald. 2005. Perception of health state and the use of vignettes to calibrate for socioeconomic status: Results of the World Health Survey in Brazil, 2003. *Cadernos de Saúde Pública* 21: S65–77.
- Ericsson, K. A., and H. A. Simon. 1980. Verbal reports as data. *Psychological Review* 87:215.
- Forsyth, B. H., and J. T. Lessler. 1991. Cognitive laboratory methods: A taxonomy. In *Measurement errors in surveys*, eds. P. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman, 393–418. New York: Wiley.
- Grol-Prokopczyk, H., J. Freese, and R. M. Hauser. 2011. Using anchoring vignettes to assess group differences in general self-rated health. *Journal of Health and Social Behavior* 52:246–61.
- Grol-Prokopczyk, H., E. Verdes-Tennant, M. McEniry, and M. Ispány. 2015. Promises and pitfalls of anchoring vignettes in health survey research. *Demography* 52:1703–28.
- Hirve, S., X. Gomez-Olive, S. Oti, C. Debpuur, S. Juvekar, S. Tollman, Y. Blomstedt, S. Wall, and N. Ng. 2013. Use of anchoring vignettes to evaluate health reporting behavior amongst adults aged 50 years and above in Africa and Asia—Testing assumptions. *Global Health Action* 6. <http://www.globalhealthaction.net/index.php/gha/article/view/21064> (accessed November 1, 2016).
- Hopkins, D. J., and G. King. 2010. Improving anchoring vignettes: Designing surveys to correct interpersonal incomparability. *Public Opinion Quarterly* 74:201.
- Johnson, T. P. 1998. Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-Nachrichten Spezial* 3:1–40.
- Johnson, T. P. 2006. Methods and frameworks for cross-cultural measurement. *Medical Care* 44:S17–20.

- Jürges, H., and J. Winter. 2013. Are anchoring vignettes ratings sensitive to vignette age and sex? *Health Economics* 22:1–13.
- Kapteyn, A., J. P. Smith, and A. Van Soest. 2007. Vignettes and self-reports of work disability in the United States and the Netherlands. *The American Economic Review* 97:461–73.
- Kapteyn, A., J. P. Smith, A. Van Soest, and H. Vonkova. 2011. Anchoring vignettes and response consistency. Labor and Population working paper series, Rand Center for the Study of Aging, Santa Monica, CA. https://www.researchgate.net/profile/Hana_Vonkova/publication/228269452_Anchoring_Vignettes_and_Response_Consistency/links/02e7e53adae1bbf2f7000000.pdf (accessed October 20, 2016).
- Keesing, R. M. 1974. Theories of culture. *Annual Review of Anthropology* 3:73–97.
- King, G., C. J. L. Murray, J. A. Salomon, and A. Tandon. 2004. Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review* 98:191–207.
- King, N. B., S. Harper, and M. E. Young. 2012. Who cares about health inequalities? Cross-country evidence from the World Health Survey. *Health Policy and Planning* 28:558–71.
- Lehman, D. R., C.-y. Chiu, and M. Schaller. 2004. Psychology and culture. *Annual Review of Psychology* 55:689–714.
- Pasick, R. J., S. L. Stewart, J. A. Bird, and C. N. D’Onofrio. 2001. Quality of data in multiethnic health surveys. *Public Health Reports* 116:223.
- Salomon, J. A., A. Tandon, and C. J. L. Murray. 2001. Using vignettes to improve cross-population comparability of health surveys: Concepts, design, and evaluation techniques. Global Programme on Evidence for Health Policy Discussion Paper No. 41. <http://www.who.int/healthinfo/paper41.pdf> (accessed October 30, 2016).
- Salomon, J. A., A. Tandon, and C. J. L. Murray. 2004. Comparability of self-rated health: Cross-sectional multi-country survey using anchoring vignettes. *British Medical Journal* 328:258.
- Su, Y. 2015. *Assessing the validity of anchoring vignettes in measuring self-rated health: A survey study in China using objective vision as a gold standard*. Boston: Department of Global Health and Population, Harvard University.
- Tourangeau, R., L. J. Rips, and K. Rasinski. 2000. *The psychology of survey response*. Cambridge: Cambridge University Press.
- Wada, K., R. Kakuma, K. Hoshi, Y. Sato, T. Hasegawa, and T. Satoh. 2011. Factors associated with preferences for health system goals in Japan: A pilot study of the World Health Survey. *Asia-Pacific Journal of Public Health* 23:721–29.
- Wand, J. 2013. Credible comparisons using interpersonally incomparable data: Nonparametric scales with anchoring vignettes. *American Journal of Political Science* 57:249–62.

-
- Wand, J., G. King, and O. Lau. 2011. Anchors: Software for anchoring vignette data. *Journal of Statistical Software* 42:1–25.
- Willis, G. B. 2005. *Cognitive interviewing: A tool for improving questionnaire design*. London: Sage.
- Willis, G. B. 2015. *Analysis of the cognitive interview in questionnaire design*. New York: Oxford University Press.
- Willis, G. B., and K. Miller. 2011. Cross-cultural cognitive interviewing seeking comparability and enhancing understanding. *Field Methods* 23:331–41.