Practical Assessment, Research & Evaluation

Is a picture is worth a thousand words? Creating effective questionnaires with pictures

Laura Reynolds-Keefer, *University of Michigan-Dearborn*
Robert Johnson, *University of South Carolina*

In developing attitudinal instruments for young children, researchers, program evaluators, and clinicians often use response scales with pictures or images (e.g., smiley faces) as anchors. This article considers highlights connections between word-based and picture based Likert scales and highlights the value in translating conventions used in word-based Likert scales to those with pictures or images.

In light of the growing effort to address "essential" elements of assessment and design for teachers and administrators, (Cizek, 1997; McMillan, 2000, 2001; Stiggins & Conklin, 1992), state educational standards and teacher preparation courses have begun attending to the importance of assessment techniques and principles. The heightened interest in assessment practices has highlighted the importance of the nature of assessment tools as used and created by practitioners (Plake & Impara, 1997). One type of assessment created and used by teachers and psychologists is a questionnaire using pictures or images instead of text as category descriptors.

Questionnaires with pictures in place of text (e.g., disagree and agree) are used by teachers, psychologists and researchers in situations where reading ability might create barriers (Zhang, Smith, Lam, Brimer & Rodriquez, 2002; Chambers & Craig, 1998).   An example of pictorial representations include the ubiquitous "smiley face" with expressions ranging from "unhappy" to "happy." A more elaborate popular example used in exploring student affect toward reading and writing uses Garfield figures showing a variety of expressions from "angry" to "happy" (McKenna & Kear, 1990; Kear, Coffman, McKenna, & Ambrosio, 2000).   There has been some exploration of the "helpfulness" of pairing images representing concrete concepts with text in the testing of young children, as well as discussion relating to the extent to which visual images or pictures promotes "meaningful response production" (Mantzicopoulos, French, & Maller, p. 1216). Some evidence suggests pairing visual images in questionnaires facilitates responses in young children (Cassidy, 1988; Eder, 1990; Harter & Pike, 1984; Measelle et al., 1998; Miller, 1985; Mize & Ladd, 1988; Verschueren, Marcoen, & Schoefs, 1996). This evidence is refuted, however, by research that this type of pairing can create confusion (Davis-Kean, 1995; Davis-Kean & Sandler, 2001; Marsh et al., 1998; Marsh et al., 2002). In the debate as to how effective this type of instrument may be, it may be useful to consider the characteristics of this type of instrument.

In considering the use of picture-based scales that allow children to report pain, Chambers and Craig (1998) noted, for example, that "children's pain ratings vary depending on the types of faces scale used, and that faces scales with smiling anchors may confound affective states with pain ratings." The relevance of the type and number of images used in questionnaires, the relationship between those images and the responses has not been thoroughly explored. Additionally, the format of questionnaires using pictures or images shows huge variance in format and style, and the "best practice" for creating such questionnaires has not been thoroughly considered. In using or creating this type of instrument, what role do the type, form, and number of image have in the selections young children make?

The purpose of this article is to present and compare the format and configurations used in several existing picture-based Likert scales, and to explore the impact differences in images and configurations might have in student responses through the lens of a small preliminary study of kindergarteners.

Characteristics of image-based Likert scales

Three characteristics unique to image-based questionnaires are the type of image, the seriation of the image, and the familiarity or complexity of the image. In considering the type of images selected for use in a questionnaire, differences in the qualities and characteristics of the images themselves can be a distinguishing factor. Many image-based questionnaires employ basic drawings that attempt in their simplicity to correspond to an emotional state. The most common example of this type of image is the use of facial drawings. These facial representations appear in two common variations in the literature, the first a cartoon "happy" face, and the second a more realistic drawing or representation of an adult human face.
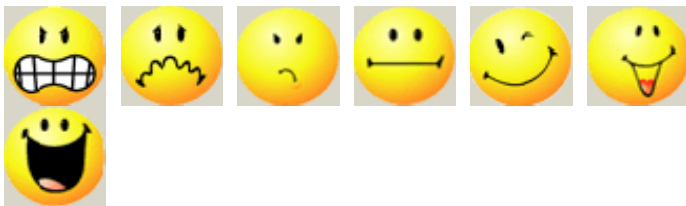
Figure 1. Teddy Bear Hospital survey using familiar "happy face" image.

Figure 2: Pediatric Pain Sourcebook of Protocols

The familiar yellow "happy" faces used in the Teddy Bear Hospital (TBH) questionnaire rely on this simple and familiar image to convey meaning, and makes no attempt at realism. The more detailed drawings used in the Pediatric Pain Sourcebook convey a greater sense of realism than the yellow cartoon faces, but this notably adult face lacks gender distinguishing features and anatomical details such as ears and hair.

In addition to the difference in image style and detail, scales using images create participant choice through differences in the set of pictures provided. Variations in the images create the differences that participants are intended to identify and use to indicate response preference. In the process of creating these scales, image variation can be necessarily subtle, and potentially difficult for younger children to discern, especially in scales of 6 or 7 images. For example, the visual scales using facial images in Figures 1 and 2 have 6 and 7 choice variants, and illustrate the subtle changes facial expressions used to indicate anchor differences, as well as the difficulty in capturing a "neutral" facial expression. The six image Pediatric Pain Scale (Figure 2) uses subtle differences between face two and three that would have to be detected and understood by the patient. In this scale, the difference in level of pain is expressed through subtle differences including the number of frown lines, the lines of the mouth, and the eyebrows. An additional concern in the representation of emotional state through images relates to the possibility of bias. Work on facial expression and interpretation indicates emotive facial expression decoding is cultural construct, and therefore ethnic and cultural bias could become an issue. The images used in questionnaires could have different meanings relative to cultural background, and these differences could impact respondent choice and the interpretation of the results (Biehl et al., 1997; Camras, Bakeman, Chen, Norris, & Cain, 2006; Huang, Nijholt, Pantic, & Pentland, 2007; Masuda, 2008; Matsumoto & Kupperbusch, 2001).

In considering the type of image and seriation, it is notable that these scales do not represent children's facial features, and the faces that are used are drawings of generic adult faces that lack any gender characteristics and some anatomical detail. These images create a scale that neutral in gender and ethnicity – essentially giving children "simple faces" to use as anchors. Figure 3 illustrates the use of a shorter scale, but with images that are of an adult face, that lack anatomical and gender features, are ethnically generic, and interpreted with text below the images.

I enjoy my reading lessons.

Figure 3: Reading Attitude Survey used with English Language Learners at the University of North Carolina

Perhaps one of the important considerations, then, is that because an attempt is commonly made to make facial images neutral, the underlying principle must be that the type of image used in the questionnaire could make a difference to the respondent. There is little literature addressing how the images used might impact respondent choice. It may be important to consider how respondents might react if the faces were female, represented a specific ethnicity, or were the faces of children.

Some Likert scales using images incorporate more complex imagery in the prompts and in the scale itself, often using cartoon animals. Two examples of the use of this type of image are the Koala Fear Questionnaire and the Elementary Reading Attitude Survey using the character Garfield. These questionnaires provide cartoon imagery of animals for children to indicate response to the question or prompt. These two scales differ slightly in that the Koala Fear Questionnaire shows children images that are designed to evoke differing levels of anxiety, and then asks children to indicate the koala bear face that most corresponds with the level of fear they feel when viewing the image. Each of the three images of the koala bears is associated with text that describe or interprets the facial expression of the image. This relies on a participant interpreting the complex event depicted in the "stem" image, and then selecting from the images to the right to indicate their personal response to the initial image. In this example, the choices have verbal descriptors ("no fear") that could be read by or to participants, however, the stem image has no text and must then be decoded and contextualized by the participant.
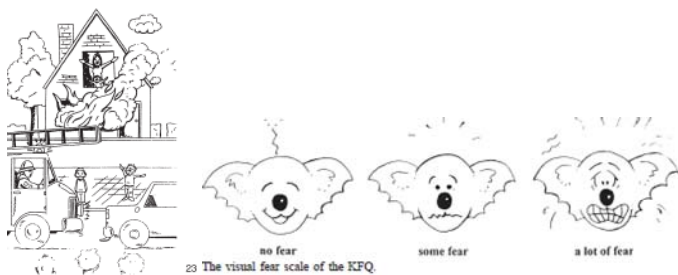

Figure 4: The Koala Fear Questionnaire (KFQ)

The Elementary Reading Attitude Survey uses the character Garfield to help students register responses to questions relating to reading (Kear, Coffman, McKenna, & Ambrosio, 2000). The four Garfield images are not associated with text descriptors, so the student has only their perception of the image's meaning. The scale relies on students differentiating between the images based on the perception of the drawing alone. Because the image is detailed, the quality of the reproduced image is also critical. Any lack of clarity in the drawings could result in an incorrect interpretation of the meaning of the images, which may be of particular concern in relation to the center two images.

With a scale using images such as Garfield, another concern may be the "attractiveness" of some of the images. If one of the images used in a questionnaire is particularly attractive, might the respondents (especially children) select it simply because they "like" that image? The image of Garfield associated with the category "Disagree" is that of an upset Garfield looking angry. If participants find this image funny, appealing, or appropriate for how that character usually behaves they may select it even if it does not reflect their response to the question or statement. Reynolds-Keefer, Johnson, Dickenson, & McFadden (2009) explored this by considering student preference by presenting students with identical questionnaires that varied only by the type of image or text used.
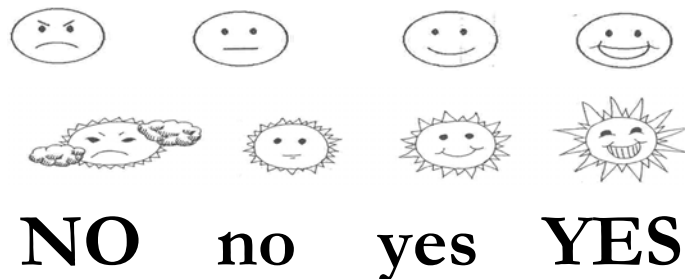

NO   no   yes   YES

Figure 5: Image choices in Reynolds-Keefer, Johnson, Dickenson, & McFadden (2009)

Although this study showed no significance in children's responses based on image alone, when gender and grade level were considered, there were statistically significant differences in responses based on the images presented. These finding highlight the possible need for further exploration of the impact different images may have for specific populations.

Pilot study

In considering the impact differences in images might have on responses, a small convenience sample of 15 kindergarten children responded to two questions relating to their feelings about reading. The sample consisted of 9 boys and 6 girls, and included two

African-American children and one Asian child. The children participating had no diagnosed or documented physical or learning disabilities, and were free to stop participating at any time. The purpose of this preliminary exploration was to explore the impact differences in images had in relation to the consistency of children's responses. Participating children were asked to answer two questions read to them each Monday for one month. The questions asked each Monday were the same, but the images provided for the children to indicate their response differed from week to week. The children participating responded to the questions during small group "writing center" work, and completed their questionnaires in less than 5 minutes each week. All 15 children completed the 4 week question cycle, despite having been given the option of not responding to the questions each week.

Students were asked two questions each week intended to reflect the child's attitude toward reading: How do you feel when you read a book at school, and how do you feel when someone reads a book to you. These questions were developed to focus on the school experience so that all students would be able to respond. Because attitude toward reading is a fairly stable construct (Guthrie, 2007, p. 283; Lewis & Teale, 1982; Smith, 1990, p. 218), any differences in response from week to week over the course of one month could indicate that the change in image used in the questionnaire might have impacted the response the child selected. This small study explores these ideas in a preliminary fashion in order to inform future more formal exploration.

Instruments

The two-item questionnaire used the images of a fairly complete human face in week 1, a generic happy face in week 2, the face of a cat in week 3, and the face of the same cat with additional graphic images in week 4. Each week, the size, font, and style of the questionnaire was identical in every way with the exception of the image used that week. The images were selected to represent a range of pleasure including: very happy, happy, unhappy, and very unhappy.

The images were selected to explore the impact of two specific issues surrounding the used of image-based Likert instruments: Does the facial realism or completeness of an image impact participant response; and do additional graphics or the interest or appeal of an image have an impact on participant response?
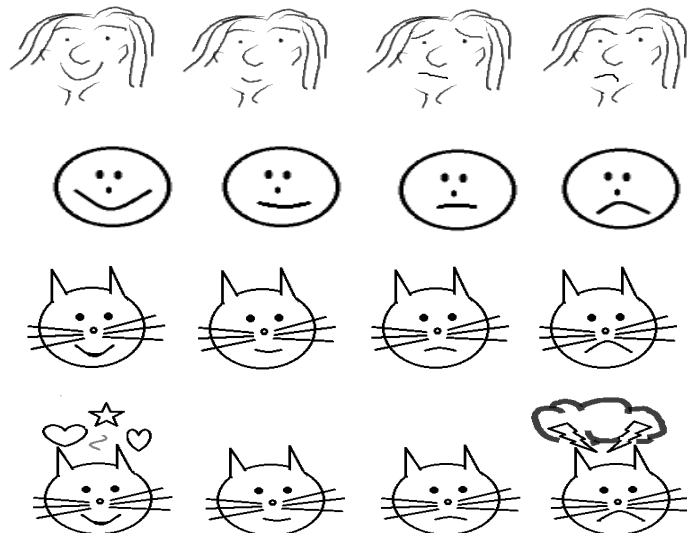

Figure 6: Images used in two-item weekly questionnaire

Results

In question 1 that asked participants to indicate how they felt when reading a book at school, 8 of the 15 children changed their responses over the course of the month. The changes occurred between weeks 1 and 2, and between weeks 3 and 4. No child altered response between weeks 2 and 3. These changes in response in all cases but 1 represented a change to reporting a more extreme attitude than what was reported previously (for example a change in response from "happy" to "very happy").

In response to question 1 that asked "How do you feel when you read a book at school" 12 of the 15 student responses were identical in weeks 1 and 2. The 3 participants that recorded different responses in week 2 did so for both questions, and all 3 changed their response from the image corresponding to "unhappy" to the image corresponding to "very unhappy". Throughout the rest of the study, these three children responded to all the questions with the most negative response, selecting the most negative images of a cat in both weeks 3 and 4.

The responses in week 3 were identical to those in week 2. In week 4, however, 3 of the 15 children answered selected images different from their responses in week 3. Of the 3 children that recorded different responses, 2 children changed from the cat indicating a "happy" to the most positive image of the cat with stars and hearts above the image. The other child that changed response between weeks 3 and 4 indicated the

moderate "unhappy" image in week 3 and the most negative image of a cat with the dark cloud and lightening above the image in week 4.

Table 1: *Responses to questionnaires by images (n=15)*

Image	Very Happy	Happy	Unhappy	Very Unhappy
Week 1 – Realistic face				
Question 1	7	2	4	2
Question 2	9	5	1	0
Week 2 – "Happy face'				
Question 1	7	2	1	5
Question 2	10	4	0	1
Week 3 – Cartoon Cat				
Question 1	7	2	1	5
Question 2	10	4	0	1
Week 4 – Cat and graphics				
Question 1	9	0	0	6
Question 2	12	1	0	2

Implications

This small study indicates that it is possible that student responses may be impacted by the types of images used in image-based Likert questionnaires. Students in this study did appear to move away from the images representing more moderate responses when the questionnaire options were less realistic and more exaggerated images. The images in the questionnaire became more exaggerated each week, with each week becoming more cartoon-like and less representative of a human face. The response pattern parallels this exaggeration, with a gradual polarizing of responses by week 4. Additionally, the similarity in the response patterns of children in weeks 2 and 3 suggest that "generic" images may illicit similar responses. Moderate responses were selected most often by participants in week 1, the week in which the image was the most "realistic" rendering of a human fact.

Very little can be generalized from this small study, however it does make the point that it is possible that the images used in questionnaires may impact the responses of participants. The polarizing of responses may be the result of the familiarity of the instrument over the four-week period, but the reasons for participants change in recorded responses remain unknown. Further research on larger groups could include follow-up interviews with participants that changed responses to probe further the impact of the change in the images used. In addition, exploring these issues with older children could provide even greater insight into the impact of the images presented as questionnaire response choices.

Conclusions

With the increasing emphasis on assessment quality, further work on image-based Likert scale may be fruitful for psychologists, teachers, and researchers (Plake & Impara, 1997; Zhang, Smith, Lam, Brimer & Rodriquez, 2002). In light of the debate regarding the significance of facila detail and cultural background (Biehl et al., 1997; Camras, Bakeman, Chen, Norris, & Cain, 2006; Huang, Nijholt, Pantic, & Pentland, 2007; Masuda, 2008; Matsumoto & Kupperbusch, 2001) as well regarding the ability of pictures to facilitate children to indicate meaningful responses (Mantzicopoulos, French, & Maller, p. 1216), it seems important to explore further the role of images in this type of frequently used assessment tool.

References

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T. & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability Data and Cross-National Differences. *Journal of Nonverbal Behavior*, 21, 3-21. doi.org/10.1023/A:1024902500935DO - 10.1023/A:1024902500935

Camras, L.A., Bakeman R., Chen Y., Norris K., & Cain T.R., (2006). Culture, ethnicity, and children's facial expressions: a study of European American, Mainland Chinese, Chinese American, and adopted Chinese girls. *Emotion*, 6(1),103-14.

Chambers, C. T., & Craig, K. D. (1998). An intrusive impact of anchors in children's faces pain scales. *Pain,* 78, 27-37.

Chan, J. (1991). Response-order effects in Likert-type scales. *Educational and Psychological Measurement, 51*, 531-540.

Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye (Ed.), Handbook of classroom assessment: Learning, adjustment, and achievement. San Diego: Academic Press.

Fowler, Jr., F.J. (2001). *Survey research methods*. Applied Social Research Methods Series, Volume 1. Sage Publications: Thousand Oaks, CA.

Frary, R. B. (1996). Hints for designing effective questionnaires. *Practical Assessment, Research & Evaluation*, 5(3). Retrieved from http://PAREonline.net/getvn.asp?v=5&n=3.

Guthrie, J.T., Hoa, A.L.W., Wigfield, A., Tonks, S., Humenick, N., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. Contemporary Educational Psycholgy, 32(3), 282-313.

Huang, T., Nijholt, A., Pantic, M., Pentland, A. & Cohn, J.(2007). Foundations of Human Computing: Facial Expression and Emotion: Artificial Intelligence for Human Computing. Springer Berlin / Heidelberg. doi.org/10.1007/978-3-540-72348-6_1DO - 10.1007/978-3-540-72348-6_1

Kear, D., Coffman, G., McKenna, M., & Ambrosio, A. (2000). Measuring attitude toward writing: A new tool for teachers. *The Reading Teacher,* 54, 10-23.

Lewis, R. & Teale, W. H. (1982), Primary School Students' Attitudes towards Reading. Journal of Research in Reading, 5: 113–122. doi: 10.1111/j.1467-9817.1982.tb00137.x

Linn, R. L., & Gronlund, N. E. (2000). Measurement and assessment in teaching (8th ed.). Upper Saddle River, N J: Prentice-Hall.

Mantzicopoulos, P., French, B.F. & Maller, S.J. (2004). Factor Structure of the Pictorial Scale of Perceived Competence and Social Acceptance With Two Pre-Elementary Samples. *Child Development*, 75(4), 1214 – 1228.

Masuda, T., Ellsworth, P., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. Journal of Personality and Social Psychology, 94(3), 365-381.DOI: 10.1037/0022-3514.94.3.365

Matsumoto, D. & Kupperbusch, C. (2001). Idiocentric and allocentric differences in emotional expression, experience, and the coherence between expression and experience - Asian *Journal of Social Psychology* 4(2) Blackwell Publishers Ltd

McKenna, M., & Kear, D. (1999). Measuring attitude toward reading: A new tool for teachers. In S. Barrentine (Ed.) *Reading assessment: Principles and practices for elementary teachers* (p. 199-214). Newark, DE: International Reading Association.

McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, 7(8). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=8.

Pediatric Pain Sourcebook of Protocols, Policies, and Pamphlets. Centre for Pediatric Pain Research, Dalhousie University and IWK, Health Centre, Halifax, Nova Scotia, Canada.

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, adjustment, and achievement* (pp. 53-68). San Diego: Academic Press.

Reynolds-Keefer, L., Johnson, R., Dickenson, T., & McFadden, L. (2009). Validity and reliability in visual Likert scales. *Studies in Learning, Evaluation, Innovation and Development, 6*(3), 15-24.

Rigby, K. (1997). *The Peer Relations Assessment Questionnaire.* Point Lonsdale: Professional Reading Guide.

Smith, M. C. (1990). A longitudinal investigation of reading attitude development from childhood to adulthood. *Journal of Educational Research,* 83(4), 215-219.

Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment.* Albany, NY: State University of New York Press, Albany.

Reading Attitude Survey used with English Language Learners (ELL) at the University of North Carolina. Retrieved from http://www.unc.edu/depts/scale/programs/arfiles/readingattitudesurvey.doc

Victorine, A., Wong, Z.Y. & Geilani, J. (2004). Teddy Bear Hospital (TBH) – reducing children's fear of doctors and hospital environment. University of Sheffield, Dr. Andrew Charters – Barnsley Primary Care Trust, Child and Adolescent Unit. Retrieved from http://www.ucl.ac.uk/teddy-hosp/tbh_barnsley_paper.pdf

Weng, L. J., & Cheng, C. P. (2000). Effects of response order on Likert-type scales. *Educational and Psychological Measurement, 60*, 908-924.

Zhang, J. J., Smith, D. W., Lam, E. T. C., Brimer, J., & Rodriquez, A. (2002). Development of an evaluation scale to measure participant perceptions of after-school enrichment programs. *Measurement in Physical Education and Exercise Science*, 6, 167-186.

Citation:

Reynolds, Laura & Robert Johnson (2011). Is a picture is worth a thousand words? Creating effective questionnaires with pictures. *Practical Assessment, Research & Evaluation*, 16(8). Available online: http://pareonline.net/getvn.asp?v=16&n=8.

Authors:

Laura Reynolds-Keefer
University of Michigan-Dearborn
lrkeefer [at] umd.umich.edu

Robert Johnson
University of South Carolina