

Reliability and Validity of the Generalized Attitude Measure and Generalized Belief Measure

James C. McCroskey

The reliability and validity of two generalized measures are evaluated based on research reported over the past four decades. These measures have been used to record attitudes and beliefs in a wide variety of communication research studies. Both the generalized attitude measure (GAM) and the generalized belief measure (GBM) were designed to be used across contexts—attitudes or beliefs concerning a wide variety topics rather than a single topic. Results of the previous research indicate that the measures are highly reliable (alpha estimates above .90) and have strong face, concurrent, and predictive validity.

Keywords: Attitude Measurement; Generalized Attitude Measure; Generalized Belief Measure; Reliability and Validity

During the 1920s and 1930s, many researchers in psychology worked to develop a variety of approaches to create instruments that could measure research participants' attitudes. The two approaches that received the most acceptance were those developed by Thurstone (1931), the method of *Equal Appearing Intervals*, and Likert (1932), the method of *Summated Ratings*. Both of these were in common usage by the 1950s. Numerous measures had been developed using these methods and found to be both reliable and valid, although there were those who were very critical of these approaches (e.g., Guttman, 1944).

Some researchers early on were experimenting with an approach known as "bipolar scaling" (McCroskey & Richmond, 1989). However, this approach was not widely used, prior to the late 1950s, when Osgood, Suci, and Tannenbaum

James C. McCroskey (DEd, Pennsylvania State University, 1966) is a Scholar in Residence in the Department of Communication Studies at the University of Alabama at Birmingham, Birmingham, AL 35294-4480, USA. Correspondence: James C. McCroskey, Department of Communication Studies, University of Alabama at Birmingham, 35294-4480; Tel: +1-205-934-3877; E-mail: email@JamesCMcCroskey.com

GAM, determined this concern was not valid in important cases (McCroskey, 1968b; McCroskey, Prichard, & Arnold, 1967). Of primary importance was the finding that single bipolar items typically do not produce interval data; however, it was found that an instrument with six bipolar items (i.e., the GAM) does generate such data. In more in-depth (though unpublished) analyses, it was also found that five-item bipolar measures also consistently produced interval data, and that four-item bipolar measures did so in 75 percent of the cases studied. Hence, it is safe to assume that data collected with the GAM will produce interval total scores.

Testing the Reliability of the GAM

Although the importance of reliability in measurement has been recognized since the earliest history of social science research, it was not common for researchers in communication (and most other social science disciplines) to report internal or test-retest reliability estimates in journal articles unless they were reporting the development of a new instrument. When McCroskey became editor of *Human Communication Research* in 1977, his call for papers was the first to mandate reports of reliability estimates for all measures in papers submitted to that journal. Such requirements now are common in most communication journals that publish quantitative research. However, in the early years after the development of the GAM, there were few reports of the internal or test-retest reliability of the instrument (e.g., Arnold, McCroskey, & Prichard, 1972; McCroskey, 1968b, 1969; McCroskey, Prichard, & Arnold, 1967; Prichard & McCroskey, 1967).

Although this tradition of non-reporting was strong in academic journals, that was not the case in doctoral dissertations. As a result, a substantial number of reliability estimates for the GAM were available in McCroskey's (1966) doctoral dissertation. For a dozen internal reliability estimates, the range was from .93 to .98. Test-retest reliability estimates for one week ranged from .84 to .96; for six weeks, the range was from .77 to .90; and for seven weeks, the range was from .72 to .89.

Testing the Validity of the GAM

The first test of validity for the GAM was examinations of face validity. A group of nine doctoral students who were well trained and experienced in measurement were asked to examine the GAM and estimate its face validity on a scale of 1–10. All estimates provided were 9 or 10. While this type of validity may not be as persuasive as more quantitative methods, they were the first indication of the quality of the GAM.

The primary validity tests of the GAM were conducted by McCroskey (1966). In his research program, Likert-type measures for attitude toward capital punishment and federal control of education were developed. A preliminary test of the concurrent validity involved Likert-type measures compared to GAM measures. The results indicated a concurrent validity correlation between the measures for capital punishment of .93, and for the federal control of education measures, the concurrent validity correlation obtained was .83. In the major study employing these measures, the

concurrent validity correlations obtained were .92 and .93 for the capital punishment measures and .81 and .86 for the federal control of education measures. These results provided very strong evidence of the concurrent validity of the measures.

Early tests of the predictive validity of the GAM yielded support for the instrument. In an international debate, these measures tapped a major change in attitude against school segregation. This was the position taken by the team identified as the winner of the debate (Prichard & McCroskey, 1967). A study of the impact of evidence as an inhibitor or facilitator of attitude change confirmed hypotheses involving the interaction and main effects. Attitude changes as measured by the GAM were consistent with those hypothesized (McCroskey, 1969). In addition to these studies, the predictive validity of the GAM was tested numerous times in McCroskey's (1966) major studies. These tests also supported the predictive validity of the GAM. Hence, all evidence from the early studies pointed to the validity of the GAM.

Initial Development of the GBM

Because it was understood that "attitude" and "belief" are different kinds of constructs, it was recognized that the GAM was not appropriate for measurement of beliefs. In addition, it was recognized that a simplified instrument that included only a few items would be desirable for circumstances where it was necessary to measure multiple beliefs. Some scholars have used single-item measures under these circumstances. Scholarly lore suggests that single items are "notoriously unreliable," even though no internal reliability can be tested with a single item, and there are numerous examples of satisfactory test-retest reliabilities for single items. However, since earlier research (McCroskey, Prichard, & Arnold, 1967) also had determined that single bipolar items did not measure equal intervals of intensity—but that six item measures did meet that criterion—it was decided to attempt to develop a measure of belief with six bipolar items.

The same basic approach to developing the GBM (also known as the Generalized Belief Scale) was employed that had been previously used to develop the GAM. That is, a large number of potentially useful items were generated and administered to numerous participants, and a factor analysis was computed in search for the six items desired. Twenty items were chosen for use and presented to more than 200 students in required classes at a very large university in the Midwest. Twenty belief statements (such as "There should be a guaranteed annual wage for all industrial employees") were randomly distributed, along with the belief items, to the student participants. Each participant responded to only one belief item.

An unrotated factor analysis was conducted. Seventeen of the items had their highest loading on the first unrotated factor. The three "bad items" were omitted, and a second factor analysis was conducted. The six items with the highest loadings (I agree-I disagree, bad idea-good idea, yes-no, false-true, right-wrong, incorrect-correct) were chosen for the final measure. The internal reliability of the instrument, based on the original data, was estimated to be .94. In subsequent research, it was determined that "good idea-bad idea" could not be applied to all kinds of belief

statements and was subsequently dropped from the measure. The deletion of this item was found to not reduce the internal reliability of the measure. A follow-up check also found that these five items yielded interval data. It also should be mentioned that one item, right-wrong, appears on both the GAM and GBM. Conversations with researchers who have used both measures indicate the belief that respondents see these terms differently (but appropriate) in the two different conditions.

It should be noted that no report of the above research was ever presented at a convention or submitted for publication. Because very vocal members of the academic community in the field at that time didn't believe such measures could be developed, the developer chose to simply use them himself and not enter a debate about whether they could be truly "general." The first time the General Attitude Measure and the General Belief Measure were published with these names was in 1996 (McCroskey & Richmond). During the intervening years, dozens of manuscripts including these measures were submitted for publication by this author. In no case did any reviewer or any editor challenge the appropriateness, reliability, or validity of these measures, which suggests they may have been totally unaware of their "general" nature.

More Recent Use of the GAM and GBM

From the late 1960s to the present, these instruments have been employed for a wide variety of research efforts. To attempt to note every case in which these instruments have been used would require more manuscript pages than the editor will permit, and as authors have not had a source to cite in many instances, it would be virtually impossible to find all of the cases where they have been used. Instead, we will provide several examples of their use and indicate the reliabilities obtained (when reported).

Use of the GAM

The most common use of the GAM in the earlier research was to measure "attitude" in a wide variety persuasion and message studies. Mehrley and McCroskey (1970) reported the first study that used fewer than six items, using only four. To measure attitude toward Brazil, they employed beneficial-harmful, good-bad, and wise-foolish (three items from the GAM), as well as valuable-worthless. To measure attitude toward the Ku Klux Klan, they employed beneficial-harmful, good-bad, fair-unfair, and wise-foolish (all items from the GAM). They obtained internal reliability estimates of .86 for Brazil and .90 for Ku Klux Klan. This is the only instance where the internal reliability for the measure has been found to be below .90.

McCroskey, Young, and Scott (1972) employed all six GAM items to measure attitude. They measured a pretest and a delayed post-test. Estimated reliabilities obtained (not reported) were .95 and .92. Arnold, McCroskey, and Prichard (1972) used the GAM to measure attitude toward 25 different topics and obtained reliability estimates averaging .95 (not reported). Wheelless and McCroskey (1973) used the GAM to measure attitude toward Brazil and obtained a reliability estimate of .94 (not reported). In all three cases, the hypotheses were supported.

In more recent research focused on organizational communication, the GAM has been used to measure subordinates' attitude (or affect) toward their supervisors and the supervisors' communication. Richmond and McCroskey (2000) obtained reliability estimates of .96 for both uses, and McCroskey and Richmond (2000) obtained reliability estimates of .95 for attitude toward the supervisor and .97 for attitude toward the supervisors' communication. Cole and McCroskey (2003) used the GAM to measure subordinates' affect for their supervisor and obtained a reliability estimate of .96. Walter, Anderson, and Martin (2005) used the six items of the GAM to measure the subordinates' "general attitude toward your communication with your supervisor" and the subordinates' "general attitude toward your supervisor." They obtained a reliability estimate of .97 for a combined score of these 12 items. In all of these studies, the hypotheses were supported. Teven, McCroskey, and Richmond (2006) used the GAM to measure subordinates' attitudes toward their supervisor and obtained a reliability estimate of .95. Again, the hypotheses were supported.

Over a period of four decades, the GAM's reliability estimates have been universally strong. Similarly, for this period of time, hypotheses have regularly been supported by data obtained through the use of the GAM. From this, it can be claimed that the GAM is a reliable and valid measure for the purposes for which it was designed.

Use of the GBM

The GBM has generally been used to measure very specific beliefs. The first published study using the GBM, for example, used as the belief to be measured "There should be a guaranteed annual wage for all industrial employees" (McCroskey & Mehrley, 1969). They used the five basic items of the GBM and also included "good idea-bad idea," which subsequently was dropped from the measure. The internal reliability estimate for the measure was .94 (with or without the sixth item). Hypotheses were supported by the study.

In another early study (McCroskey & Wright, 1971), the items from the GBM were employed to measure the belief that "appropriations for the United States space program should be substantially reduced." The obtained reliability was .94 (not reported). In another early study, Luchok and McCroskey (1978) used the GBM items to measure the belief that "the United States should adopt a new system for financing health care for all citizens." They obtained a reliability estimate of .94 (not reported). In both cases, hypotheses were supported.

One of the most extended uses of the items of the GBM was the result of their use as a part of the original Generalized Teacher Immediacy scale (Andersen, 1978). Andersen employed the GBM items as the base of her scale and added four other items. Obtained internal reliabilities for this scale in her research were .96 and .97. She also obtained a test-retest reliability of .81. This scale was used for a number of years, until researchers developed a preference for behavior-based measures of immediacy. During its use, the reliabilities reported continued to be high, and hypotheses were regularly supported in that research.

In more recent research, as was the case with the GAM, the GBM has been employed in organizational communication research. Richmond and McCroskey (2000) and McCroskey, Richmond, Johnson, and Smith (2004) have used the GBM to measure the belief that "I am very satisfied with my current job." Estimated internal reliability in the first paper was .97. The second paper reported two studies. The estimated internal reliabilities in these studies were .96 and .97. McCroskey, McCroskey, and Richmond (2005) also used the GBM to measure subordinate job satisfaction and obtained an internal reliability estimate of .95, and Teven, McCroskey, & Richmond (2006) also used it for this purpose and obtained a reliability estimate of .97. Hypotheses in all of these studies were confirmed.

The GBM has also been employed in some interpersonal communication studies. McCroskey and Teven (1999) adapted the GBM to measure liking and believability. The obtained internal reliability estimates were .95 for both beliefs. Coie and McCroskey (2003) use the same procedure with the GBM and obtained a reliability estimate of .96. Again, hypotheses in all of these studies were confirmed.

Over the period of almost four decades, the GBM's reliability estimates have been universally strong. Similarly, for this period of time, hypotheses have regularly been supported by data obtained through the use of the GBM. From this it can be claimed that the GBM is a reliable and valid measure for the purposes for which it was designed.

Discussion

While the GAM and GBM can be used for a wide variety of purposes, different attitudes, and belief targets, it must be recognized that neither is a "universal" measure. For example, only single dimensional attitude or belief targets are amenable to their use. Belief statements such as "I believe that Mary and Tom should be promoted" would require that separate GBMs would need to be administered for both Mary and Tom. Similarly many constructs are multidimensional, such as source credibility, interpersonal attraction, and homophily. The dimensions of these constructs would need to be measured separately. It should also be recognized that attitude toward "France" and "Germany" would need to be measured separately by individual GAMs for the two countries.

It should also be recognized that the use of either the GAM or GBM has limited utility for use in intercultural studies. Translating bipolar items from English to any other language can be very difficult, if not impossible. Terms that are bipolar in one language or culture frequently are not bipolar in another. Researchers should consider developing language- and/or culture-specific bipolar instruments rather than attempting to translate the English versions to any other language. Similarly, researchers need to be sensitive to differences in cultures between groups that speak what generally is thought to be the same language, such as Spanish in Puerto Rico and Mexico.

Contrary to traditional beliefs, it is clear that generalizable measures of attitude and belief, and possibly other constructs, can be developed. However, all such

measures are limited to the specific construct being studied. Yet some constructs are the product of a single or limited number of cultures. One should not attempt to measure a construct in a culture that does not have that construct or where that construct is very different between cultures. This is true for the GAM, the GBM, or any other measure.

References

- Andersen, J. F. (1978). The relationship between teacher immediacy and teacher effectiveness. Unpublished dissertation, West Virginia University.
- Arnold, W. E., McCroskey, J. C., & Prichard, S. V. O. (1972). The intensity component of semantic differential scores for measuring attitude. *Western Speech*, 36, 251-268.
- Cole, J. G., & McCroskey, J. C. (2003). The association of perceived communication apprehension, shyness, and verbal aggression with perceptions of source credibility and affect in organizational and interpersonal contexts. *Communication Quarterly*, 51, 101-110.
- Cronkhite, G. (1976). Effects of rater-concept scale interactions and use of different factoring procedures upon evaluative factor structure. *Human Communication Research*, 2, 316-329.
- Darnell, D. K. (1966). Concept scale interaction in the semantic differential. *Journal of Communication*, 16, 104-115.
- Diab, L. N. (1965). Some limitations of existing scales in the measurement of social attitudes. *Psychological Reports*, 17, 427-430.
- Guttman, L. L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9, 139-150.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-55.
- Luchok, J. A., & McCroskey, J. C. (1978). The effect of quality of evidence on attitude change and source credibility. *Southern Speech Communication Journal*, 43, 371-383.
- McCroskey, J. C. (1966). Experimental studies of the effects of ethos and evidence in persuasive communication. Doctoral dissertation, Pennsylvania State University.
- McCroskey, J. C. (1968a). *An introduction to rhetorical communication: The theory and practice of public speaking*. Englewood Cliffs, NJ: Prentice-Hall.
- McCroskey, J. C. (1968b). Latitude of acceptance and the semantic differential. *Journal of Social Psychology*, 74, 127-132.
- McCroskey, J. C. (1969). The effects of evidence as an inhibitor of counter-persuasion. *Speech Monographs*, 37, 188-193.
- McCroskey, J. C., & Mehrley, R. S. (1969). The effects of disorganization and nonfluency on attitude change and source credibility. *Speech Monographs*, 36, 13-21.
- McCroskey, J. C., Prichard, S. V. O., & Arnold, W. E. (1967). Attitude intensity and the neutral point on semantic differential scales. *Public Opinion Quarterly*, 31, 642-645.
- McCroskey, J. C., & Richmond, V. P. (1989). Bipolar scales. In P. Emmert & L. L. Barker (Eds.), *Measurement of communication behavior* (pp. 154-167). New York: Longman.
- McCroskey, J. C., & Richmond, V. P. (1996). *Fundamentals of human communication*. Prospect Heights, Ill.: Waveland Press.
- McCroskey, J. C., & Richmond, V. P. (2000). Applying reciprocity and accommodation theories to supervisor/subordinate communication. *Journal of Applied Communication Research*, 28, 278-289.
- McCroskey, J. C., Richmond, V. P., Johnson, A. D., & Smith, H. T. (2004). Organizational orientations theory and measurement: Development of measures and preliminary investigations. *Communication Quarterly*, 52, 1-14.
- McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communication Monographs*, 66, 90-103.

- McCroskey, J. C., & Wright, D. W. (1971). A comparison of the effects of punishment oriented and reward oriented messages in persuasive communication. *Journal of Communication*, 21, 83-93.
- McCroskey, J. C., Young, T. J., & Scott, M. D. (1972). The effects of message sidedness on inoculation against counter persuasion in small group communication. *Speech Monographs*, 39, 205-212.
- McCroskey, J. C., McCroskey, J. C., & Richmond, V. P. (2005). Applying organizational orientations theory to employees of profit and non profit organizations. *Communication Quarterly*, 53, 21-40.
- Mehrely, R. S., & McCroskey, J. C. (1970). Opinionated statements and attitude intensity as predictors of attitude change and source credibility. *Speech Monographs*, 37, 47-52.
- Osgood, C. E., Suci, G. I., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, Ill.: University of Illinois Press.
- Prichard, S. V. O., & McCroskey, J. C. (1967). An empirical look at an international debate. *Journal of the American Forensic Association*, 4, 21-25.
- Richmond, V. P., & McCroskey, J. C. (2000). The impact of supervisor and subordinate immediacy on relational and organizational outcomes. *Communication Monographs*, 67, 85-95.
- Teven, J. L., McCroskey, J. C., & Richmond, V. P. (2006). Communication correlates of perceived Machiavellianism of supervisors: Communication orientations and outcomes. *Communication Quarterly*, 54, 127-142.
- Thurstone, L. L. (1931). The measurement of social attitudes. *Journal of Abnormal and Social Psychology*, 26, 239-269.
- Walter, H. L., Anderson, C. M., & Martin, M. M. (2005). How subordinates' Machiavellianism and motives relate to satisfaction with superiors. *Communication Quarterly*, 53, 57-70.
- Wheless, L. R., & McCroskey, J. C. (1973). The effects of selected syntactical choices on source credibility, attitude, behavior, and perception of message. *Southern Speech Communication Journal*, 38, 213-222.