Does Religious Faith Improve Test Performance?

Mara S. Aruguete  
*Lincoln University*

Alan K. Goodboy  
*Bloomsburg University*

William J. Jenkins  
*Mercer University*

Daniel H. Mansson  
*Penn State University-Hazleton*

Lynn E. McCutcheon  
*NAJP*

In two related studies we sought to determine if reminding persons of one’s religious faith would cue an increase in self-efficacy and perseverance that would lead to better performance on an intellectual task. In experiment one college students were randomly assigned to a group that filled out a religious orientation scale or one that filled out a distracter scale. Both groups responded to a self-efficacy scale, then, took a logical reasoning test. Results showed that completing the religious orientation scale did not influence self-efficacy, perseverance, or logical reasoning. In experiment two we used a stronger religious cue (prayer). Similar to experiment one, we found no significant differences among those who did or did not pray on either self-efficacy, perseverance, or logical reasoning. Results were discussed from the framework of recent research showing that superstitions “work” by raising task-related self-efficacy, which in turn, leads to greater perseverance and improved performance.

There are several studies in the extant literature showing that brief contact with stimuli that are reliably associated with certain kinds of behavior can evoke or cue that behavior. For example, Berkowitz and LePage (1967) found that participants who momentarily handled a gun subsequently behaved more aggressively than those who handled an object not associated with violence. More recently, Klinesmith, Kasser, and McAndrew (2006) found that contact with guns served as a cue for aggression, and Mast and McAndrew (2011) showed that merely listening to violent lyrics embedded in heavy metal music caused males to behave more aggressively than males who listened to non-violent heavy metal music or no music at all. More relevant to the present study is the finding that Biblical passages that described violence triggered more violence in college students as compared to a control group told...
that the passages came from an ancient scroll (Bushman, Ridge, Das, Key, & Busath, 2007).

Are there cues that prime behaviors other than violence? A recent paper showed that belief in superstition can indirectly improve performance on several tasks (Damisch, Stoberock, & Mussweiler, 2010). The authors reported four experiments in which they identified and activated common superstitious cues (e.g., keeping one’s fingers crossed, lucky charm). In all four, those who had been randomly assigned to the group in which a superstition had been activated performed better on a motor skills task, a memory task, or an anagram game as compared to a control group. In two of the four experiments the authors showed that superstitions “work” by enhancing task-related self-efficacy, which in turn improves performance, probably because those with higher self-efficacy persist longer at the task.

Many studies have been designed to determine if religious faith confers benefits to true believers (see Seeman, Dubin, & Seeman, 2003, for a review). Seeman et al. (2003) concluded that the evidence does link Judeo-Christian religious practices to healthier blood pressure and immune functioning. Graham and Haidt (2010) argue that the health and happiness benefits enjoyed by religious people stem from participation in a religious congregation and a sense of belonging to a religious community. Others provide support for the argument that religious faith alleviates the anxiety that originates from knowing that we humans will eventually die (Jonas & Fischer, 2006; Vail, Rothschild, Weise, Solomon, Pyszczynski, & Greenberg, 2010).

Does religious faith “work” in the same way that superstitions do? Superstitions have been described as unfounded beliefs (Damisch et al., 2010), as have religious beliefs (Dawkins, 2006). People are most likely to engage in superstitious (Whitson & Galinsky, 2008) and religious behavior (Kay, Gaucher, McGregor, & Nash, 2010) when they experience low levels of perceived control.

We wondered if religious faith might also confer performance benefits. Specifically, we hypothesized that religiosity, cued by responding to a scale measuring religious beliefs, would enhance task-related self-efficacy and performance in much the same way that a primed superstition (e.g., a lucky rabbit’s foot) did in the Damisch et al. (2010) study.

**EXPERIMENT 1**

**Method**

Participants. The sample consisted of 141 college students (81 females, 52 males, and 8 with missing gender) from one private and three public universities, (two located in PA, one in MO, and one in GA)
ranging in age from 18-50 years ($M = 20.48; SD = 3.97$). The private university has historical ties to a large Protestant denomination, but is no longer officially connected with it. Participants were randomly assigned to an experimental ($n = 70$) or control ($n = 71$) group.

Materials. To manipulate the variable of religious faith, we used a priming technique. The experimental group filled out an eight-item version of Maltby’s (1999) 12-item Age-Universal I-E Scale. We used this shorter version because of time constraints due to other measures we were using. This scale has good psychometric properties, and is designed to measure three factors of religious orientation. We chose four of the six items from the Intrinsic subscale, examples of which include “I enjoy reading about my religion” and “It is important to me to spend time in private thought and prayer.” Persons who score high on Intrinsic (In) are those who view religion as important and deeply personal. We selected two of the three items from the Extrinsic-Personal (E-P) subscale, “Prayer is for peace and happiness” and “I pray mainly to gain relief and protection.” Persons who score high on E-P are those who view religion as a means to gain personal comfort and protection. We also chose two of the three items from the Extrinsic-Social (E-S) subscale, “I go to church mainly because I enjoy seeing people I know there” and “I go to church mostly to spend time with my friends.” Persons who score high on E-S are those who view religion as a way of gaining and maintaining friendships. All eight items were chosen because they had the highest factor loadings on their respective subscales (Maltby, 1999). Cronbach’s alpha reliability for the items we chose was .80.

Our control group filled out an eight-item “Personality Scale” meant to serve as a control for the possibility that filling out any sort of measure might alter responses on the dependent variables. These “personality” items were chosen and adapted from a wide variety of measures found in Measures of Personality and Social Psychological Attitudes (Robinson, Shaver, & Wrightsman, 1991). Examples include “If I have a problem, I prefer to work it out myself,” “My sleep is restless unless I am very tired,” and “Any person can find a job right now if enough effort is made.” The intent was to present no discernable pattern to the scale, to negate the possibility that a particular type of personality scale might serve as a cue that would inadvertently impact self-efficacy or the dependent variable, and so that participants would not be able to correctly guess that our intent was to merely distract them.

The eight items selected from the Age-Universal I-E Scale consisted of a total of 81 words. The eight items chosen for the “Personality Scale” also consisted of a total of 81 words. This was done deliberately to rule out the possibility that scale length might influence scores on the
dependent variable. Both scales used a seven-point Likert scale, with “strongly agree” = 7 and “strongly disagree” = 1, as the anchors.

All participants filled out a five-item Likert scale intended to measure self-efficacy. This scale used a nine-point Likert scale, with “strongly agree” = 9 and “strongly disagree” = 1, as the anchors. A German version of this scale was used successfully (Cronbach’s alpha .93 & .92) in the superstition study done by Damisch, et al. (2010). Damisch translated this scale into English, and we modified it slightly to read better. Cronbach’s alpha reliability for our sample was .76. Sample items include “I believe that I will master the reasoning skills test quite well,” “I look forward to the reasoning skills test with a good feeling,” and “I trust in my abilities to reason well.” On one of the five items (“How well do you think you will do...”) the anchors were “Not well at all” = 9 and “Very well” = 1 (reverse scored). The wording of the items reflects Bandura’s thinking that the measurement of self-efficacy should be task-specific (1997; 2006; see Smith, Kass, Rotunda, & Schneider, 2006, for more on the measurement of self-efficacy).

We included two additional items to measure motivation of participants, both using a nine-point Likert scale, with “strongly agree” = 9 and “strongly disagree” = 1, as the anchors. The first, “My performance on this test is an important indication of how well I am likely to do in graduate school, medical or dental school, or law school” was intended as a check to make sure that participants believed that the test they were about to take really was important. The second, “I am considering the possibility of going on to either graduate school, medical or dental school, or law school” was intended as a check to see if there was sufficient motivation to perform well on the reasoning skills test.

The reasoning skills test, which we named the Advanced Reasoning Skills Test (ARST), consisted of 15 items from the Scottsdale Test of Critical Thinking (J. P. Ricker, personal communication, Jan. 12, 2011), six items that we developed ourselves, two items adapted from The Elements of Reasoning (Conway & Munson, 2000), and two modified items from the California Critical Thinking Skills Test (Facione, 1990). We measured two dependent variables related to this test: 1) Time-on-task (number of seconds needed to complete the test) and 2) Test score (number of items answered correctly on the test). A few sample items can be found in Appendix A.

Procedure. Students were recruited from their respective universities to participate in a study of reasoning skills. Students participated in groups ranging in size from four to 22. Upon arrival at the testing site, participants were randomly given a questionnaire consisting of either the eight-item version of the Age-Universal I-E Scale or the eight-item “Personality Scale,” although both were labeled as “Personality Scale.”
Otherwise, the questionnaires were identical. Participants were informed in print that the study was about the “relationship between personality traits, attitudes, and reasoning skills,” and that it would take about 40 minutes. Upon finishing the eight items, participants were encouraged to reflect on their answers for a moment before turning the page. To minimize the likelihood of differential impact on the dependent variable caused by differences among the experimenters, all of the important instructions were written on the pages of the questionnaire.

The next page was labeled “Advanced Reasoning Skills Test.” It contained information about the test, namely that it was difficult, that it consisted of 25 items, and that 35 minutes would be allotted for completing it. Furthermore, participants read that “tests like this one are often used to determine who gains admittance to graduate, medical or law school.” The purpose of this was to elevate the arousal level of participants so they would be motivated to perform well. Participants then responded to the five self-efficacy items and the two items about graduate/medical school. The experimenter waited until everyone had finished these seven items, then, instructed everyone to begin. The experimenter remained in the testing room the entire time, announcing at the 15-minute-mark that “You should be nearing completion,” and at the 20-minute mark “You should be working on the last few items.” In spite of the stated time limit and the two verbal warnings, we did not enforce a time limit. Instead, we recorded the time at which each participant finished. We did this because Damisch et al. (2010) found that participants in their superstition group persevered at the task longer, thus at least partially accounting for their superior performance.

Results and Discussion

We hypothesized that religiosity, primed by responding to a scale measuring religious beliefs, would enhance self-efficacy and performance on a logical reasoning test. We tested this hypothesis using independent groups t-tests comparing the self-efficacy, time on task, and scores on the logical reasoning test among experimental and control groups. Tests on these dependent variables were not statistically significant. Scores on self-efficacy (experimental group $M = 31.69, SD = 5.55$; control group $M = 32.49, SD = 5.24$; $t_{139} = -.89, p = .38$), time on task (experimental group mean = 1082 sec, $SD = 267$; control group mean = 1128 sec, $SD = 2.46$; $t_{139} = -1.07, p = .29$), and logical reasoning (experimental group $M = 16.43, SD = 4.49$; control group $M = 16.11, SD = 4.59$; $t_{139} = .41, p = .68$), were similar in both groups.

The only measure that significantly differed between groups was a single item measure (“My performance on this test is an important indication of how well I am likely to do in graduate school, medical or
dental school, or law school"). Apparently, being primed with religion ($M = 4.86, SD = 2.32$) made participants perceive the test to be a more serious indicator of future performance than if they were not primed ($M = 3.86, SD = 1.91; t_{139} = 2.79, p < .01$).

We suspected that our lack of statistically significant results may have been influenced by including participants in the experimental group who were not religious. To test this possibility, we excluded data from 18 participants who scored more than one standard deviation below the mean on the intrinsic religion subscale of the Age-Universal I-E Scale. Even after excluding these participants, we still failed to find statistically significant differences between groups on self-efficacy ($t_{121} = -0.80, p = .43$), time on task ($t_{121} = -0.73, p = .46$), or logical reasoning measures ($t_{121} = -0.22, p = .83$).

There are two main explanations for why our religious manipulation did not result in significant changes in performance. One possibility is that the manipulation was not strong enough. Answering questions about religious beliefs may have increased the salience of religion. However, perhaps the religious manipulation needs to be more expressly tied to the performance task. For example, Damisch et al. (2010) plainly instructed participants to bring a “lucky charm” with them on a day in which they were to take a memory test. In a religious manipulation, this procedure might be analogous to having participants write out a prayer they could use to pray for superior performance on an upcoming test. Carried a step farther, participants could be asked to actually petition God using that prayer. Another possibility for the non-significant results is that religion does not influence task performance in the same way as superstitious beliefs such as lucky charms. Experiment 2 further tests how religion might influence performance by strengthening the religious manipulation.

**EXPERIMENT 2**

**Method**

Participants. The total sample consisted of 139 college students (83 females, 41 males, and 15 with missing gender) from the same four universities as in Experiment 1, ranging in age from 18 – 52 years ($M = 20.48; SD = 4.43$). Participants were randomly assigned to experimental group one, “write prayer only” ($n = 40$), experimental group two, “write and pray” ($n = 45$), or a control ($n = 54$) group. Both experimental groups were then asked to write out a 15-to-50 word prayer “to respectfully ask the God of your religion to give you the help you need to earn a high score on the test of reasoning ability.” Additionally, the “write and pray” group was asked to silently offer this prayer to the God of their religion. The control group was given a neutral writing task: “Please take a few
minutes to write out a 15-to-50 word self-description that you could use to help someone understand very quickly what your personality is like."

Except for the differences noted above, the questionnaires were identical for all three groups.

Materials. Both experimental groups completed four items from Maltby’s (1999) Age-Universal I-E Scale. We chose the same four items used previously from the Intrinsic subscale. Persons who score high on Intrinsic (In) are those who view religion as important and deeply personal. To this we added a fifth item, “I believe that God will help me earn a high score on the test I am about to take.” The purpose of this fifth item was to identify religious believers who also believe that God intervenes in human affairs.

Our control group completed the first five items from the “Personality Scale” used in the first experiment. As in experiment one, our purpose was to present a time-wise comparable distraction with no discernable pattern to the scale.

The four items on the Age-Universal I-E Scale plus the “God will help me” item consisted of a total of 61 words. These five items were presented to both experimental groups. The five items chosen for the “Personality Scale” consisted of a total of 55 words. Both scales used a seven-point Likert scale, as before.

All participants completed the same five-item Likert scale intended to measure self-efficacy that was used in Experiment One. Reliability on this five-item scale was initially low (.62), necessitating that we remove one item. On the resulting 4-item self-efficacy measure, Cronbach’s alpha was .86.

We included the same two items from Experiment One about performance as an indication of success in graduate school, intended as a check to make sure that participants believed that the test they were taking was important. The second item, considering the possibility of going on to advanced schooling, was intended as a check to see if there was sufficient motivation to perform well on the ARST.

Previous studies (Hogg, et al., 2010; Kay, Gaucher, McGregor & Nash, 2010; McGregor, Nash, & Prentice, 2010) suggest that religious faith may be useful in reducing anxiety. We selected the first five items from the “state” part of the State-Trait Anxiety Inventory (Spielberger, 1983), because we expected that the thought of taking an important test would temporarily raise anxiety level in the control group as compared to the two experimental groups. Cronbach’s alpha reliability on the abbreviated anxiety scale was .84.

We used the same reasoning skills test, the Advanced Reasoning Skills Test (ARST) as before. We also measured Time-on-task (perseverance) and Test score, as before.
Immediately following the ARST we asked “In your opinion, what is the main question this research addresses?” and “Did you pray to God to help you on this test, before or during the test?” The first item was used to eliminate participants who correctly guessed the main hypothesis; the second as a manipulation check to see if participants in the experimental group(s) did pray, and to determine that participants in the control group did not. We initially collected data on 168 participants. We excluded 29 participants from the analysis because they either guessed the hypothesis correctly, prayed in the control condition, did not pray in the experimental groups, indicated that their performance on the ARST was not important, or were not considering graduate education.

Procedure. Students participated in groups ranging in size from 1 to 28. Upon arrival at the testing site, participants were randomly assigned to one of the three groups described above.

Results and discussion

We hypothesized that offering a prayer would lower anxiety and enhance self-efficacy and performance on a logical reasoning test. We tested this hypothesis using one-way ANOVAs comparing the self-efficacy, time on task, anxiety, and scores on the logical reasoning test among the two experimental groups and the control group. There were no statistically significant differences between groups. Scores on self-efficacy ($M = 6.60, SD = 1.20; F_{132} = .61, p = .55$), time on task ($M = 1157.38$ sec, $SD = 320.94; F_{132} = 1.19, p = .31$), anxiety ($M = 1.77, SD = .66; F_{132} = .21, p = .81$) and logical reasoning ($M =16.33, SD = 3.81; F_{131} = 1.72, p = .18$), were similar in control and experimental groups.

We wondered whether our lack of statistically significant results may have been influenced by participants in the experimental groups who were not religious, or who indicated that they did not believe in a God that intervened in human affairs. To test these possibilities, we reanalyzed data excluding experimental groups participants who scored more than one standard deviation below the mean on the intrinsic religion subscale of the Age-Universal I-E Scale. This analysis yielded no significant effects on self-efficacy ($F_{117} = .22, p = .80$), time on task ($F_{117} = .1.35, p = .26$), anxiety ($F_{117} = .58, p = .56$) or logical reasoning measures ($F_{117} =1.11, p = .33$). We also reanalyzed data excluding the 39 participants who indicated that they did not believe in a God that intervened in human affairs. Even after excluding these participants, we still failed to find statistically significant differences between groups on self-efficacy ($F_{97} = .67, p = .51$), time on task ($F_{97} =.05, p = .95$), anxiety ($F_{97} = .28, p = .76$) or logical reasoning measures ($F_{96} =.30, p = .74$).

A reviewer raised the possibility that the control task of writing about one’s personality in our second experiment might have yielded
spontaneous comments about religious faith, which might have served as religious cues. We think this is unlikely, given that deliberate attempts to cue religious faith failed in the first experiment. Nevertheless, we checked what control group participants wrote in experiment two, and found that none of them spontaneously mentioned God or religious faith.

GENERAL DISCUSSION

In two related experiments we sought to determine if religious priming could affect performance like superstition affects performance, namely by increasing self-efficacy and perseverance. We found no significant differences in self-efficacy, anxiety, perseverance, or performance on a test of logical ability in either experiment, in spite of several manipulation checks and the use of a strong religious cue in the second experiment.

When a constructive replication fails to yield results conceptually similar to the experimental model on which it was based, a number of possible explanations arise. The participants in the four experiments conducted by Damisch et al. (2010) were students at a German university, raising the possibility of cross-cultural differences in the way persons perceive superstitious beliefs.

A second possibility is that our sample consisted of too many persons who were not “strong” believers in the possibility that a supreme being intervenes in relatively mundane matters. This possibility seems more likely given that our experimental groups did not experience higher self-efficacy or perseverance scores in either experiment. We also note that had we used more stringent criteria on the religion measures, we would have lost many more participants and thus experienced a drop in power. College students confront challenges to their beliefs during the four-year period when many of them are exposed to new ideas, and that often results in a decline in the importance of religious values (Feldman & Newcomb, 1969; Lefkowitz, 2005).

One limitation of our study is that we only have self-report data on whether participants actually prayed or not, and no certain way of knowing if they actually believed their prayers for help on the test would be answered.

Bell (2002) reported that 39 out of 43 studies show a negative relationship between IQ and religious belief. Killeen, Wildman and Wildman (1974) found a negative relationship between IQ and superstitiousness. Alcock and Otis (1980) found that belief in paranormal phenomena correlated negatively with critical thinking ability. With the wisdom of hindsight, perhaps a less educated sample consisting of highly intrinsically motivated religious persons who pray every day and strongly
believe that their God grants every request, might have yielded positive results.

Yet another possible explanation for our failure to constructively replicate Damisch et al. (2010) is that there are limits to the generalizability of the performance benefits of superstitions. For reasons as yet unknown, perhaps our hypothesis that religious primes activate self-efficacy and perseverance, resulting in superior test performance, fall outside those limits. We hope (but not necessarily pray) that further research will clarify the issues our research raised.

REFERENCES
Killeen, P. R., Wildman, R. W., & Wildman, R. W. II. (1974). Superstitiousness and intelligence. *Psychological Reports*, 34, 1158. DOI: 10.2466/pr0.1974.34.3c.1158


**APPENDIX A**

Examples of items found on the ARST

1. Amiya has four cards, each with a different symbol on it: a star, a triangle, a square, and a circle. What is the probability that, if the cards are well shuffled and placed in a pile face down, the top card will be the triangle?
   A. 3 out of 4               C. 4 out of 4
   B. 1 out of 4               D. 2 out of 4

2. “Not all the men are qualified to serve,” expresses basically the same idea as:
   A. None of the men are qualified to serve.
   B. At least one man is not qualified to serve.
C. Someone qualified to serve is not a man.
D. All men are not qualified to serve.

3. If Bryan is taller than Kevin, and Kevin is taller than Garrett, which of the following can Sara conclude with absolute certainty?
   A. Bryan is taller than Garrett.
   B. Bryan is taller than Sara.
   C. Kevin is taller than Bryan.
   D. Kevin is taller than Sara.

4. “All Mercurians tell lies,” means the same thing as:
   A. If anyone is Mercurian, then that person is a liar.
   B. If anyone is a liar, then that person is a Mercurian.
   C. There is at least one person who is a Mercurian who lies.
   D. People don’t lie unless they are Mercurian.

5. Consider this argument: “Person A is shorter than person R. Person Y is shorter than person A, but person M is shorter than person Y. Therefore, person Y is shorter than person J.” What information must be added to require that the conclusion be true, assuming that all premises are true?
   A. Person A is taller than J.
   B. Person R is taller than J.
   C. Person J is taller than A.
   D. Person J is taller than M.

6. Consider this argument: “All the relatives of the deceased were at the funeral. Therefore, there is no denying that everyone at the funeral was a relative of the deceased.” Assuming the first statement is true, what is the most reasonable conclusion that we can draw from the second statement?
   A. Everyone at the funeral was a relative of the deceased.
   B. The second statement is definitely false.
   C. The second statement is definitely true.
   D. The second statement is probably false.