An Initial Investigation into the Role of Stereotype Threat in the Test Performance of College Students with Learning Disabilities

Alison L. May
Northwestern University

C. Addison Stone
University of Michigan

Abstract

In a quasi-experimental evaluation of the possible role of stereotype threat in the academic performance of college students with learning disabilities (LD), students with (N = 29) and without (N = 62) identified LD took a simulated Verbal GRE® task in one of two conditions modeled after those used in past stereotype threat (ST) research. The task was presented as either a measure of verbal reasoning and reading abilities (ST condition) or as a measure of problem-solving style (Reduced-Threat condition; [RT]). The traditional ST-related performance differences in accuracy were not found in the sample of undergraduates with LD; however, marginally significant differences in the efficiency with which students with LD completed test items in the RT condition relative to the ST condition suggest the need for follow-up research, given the frequency with which the accommodation of extra time on exams is granted to and is used by undergraduates with LD. Potential limitations due to size and representativeness of the sample are addressed.

Keywords: Disability stereotypes, test performance, processing speed, ability assessments

Although school-age children with learning disabilities (LD) have traditionally received far more attention from researchers and policy makers, numerous studies note a dramatic increase in the number of individuals with LD attending college over the past 10 to 15 years (Orr & Hammig, 2009; Sparks & Lovett, 2009a). Reliable incidence figures are difficult to obtain and vary across institutions, but it appears that students with documented LD constitute on average between 2.4 - 3% of full-time freshman undergraduates in the United States (DaDeppo, 2009).

Unfortunately, as is the case for younger students with LD, college students with LD struggle more than their peers. For instance, studies have documented that these students experience more test anxiety (Davis, Nida, Zlomke, & Nebel-Schwalm, 2009; Holzer, Madaus, Bray, & Kehle, 2009) and tend to earn lower grades and fewer honors than undergraduates without LD (DaDeppo, 2009; Sparks & Lovett, 2009a). Moreover, undergraduates with LD experience prejudice from their peers and even from college professors (Hartman-Hall & Haaga, 2002; May & Stone, 2010; Murray, Wren, & Keys, 2008), which may result in professors failing to accommodate students with LD or students not approaching disability services offices or their professors in the first place. These views may be explicitly or implicitly communicated to the undergraduates with LD themselves (Quinlan, Bates, & Angell, 2012), and thus may be damaging, given the significant impact of teacher perceptions of students with LD in elementary and secondary education (DeSimone & Parmar, 2006).

In contrast to these stereotypes and in accordance with the definition of LD, college students with LD have at least average intelligence (Sparks & Lovett, 2009a). In the case of the more successful subset of

students with LD who have progressed to the post-secondary level and gained admission to highly competitive universities, one could reasonably argue that above-average conceptual and problem-solving skills are necessary in order to compensate for the students' defining skill deficits. With appropriate accommodations, students with LD should therefore be experiencing more success in their undergraduate careers than they presently are. Specifically, recent data suggest that these students take longer to graduate and that they are more likely to transfer to different colleges or to drop out altogether, even as early as their first year (DaDeppo, 2009; Murray, Goldstein, Nourse, & Edgar, 2000; Orr & Hammig, 2009). For instance, Murray et al. (2000) found that undergraduates with LD had a 23% lower graduation rate than their peers without LD. Studies also report undergraduates with LD to earn a grade-point-average (GPA) as much as one-half of a point below their non-disabled peers (DaDeppo, 2009; Zurcher & Bryant, 2001). One might question, therefore, how much of the lower achievement by undergraduates with LD is attributable to their learning disabilities, and how much is a result of external factors such as prejudice and lower expectations for these students.

This account is unfortunately very similar to the case of many students of color at the college level. For example, Osborne (2007) found high-performing African American college students to be particularly susceptible to various measures of test anxiety. Furthermore, Steele (1999) noted that the dropout rate for African American college students is 20 to 25% higher than that for Caucasians, and that those who complete college generally earn a GPA that is two-thirds of a grade point below that of Caucasians. These disparities persist for even middle-class African American students, suggesting that lack of social or economic opportunities is not a sufficient explanation for these students' failures.

Stereotype Threat

In an attempt to explain the disproportionate number of failures experienced by African American undergraduates, Claude Steele, Joshua Aronson, Steven Spencer, their associates, and many other research groups in the last 15 years have provided evidence for a theoretical construct they call "stereotype threat" (Beasley & Fischer, 2012; Ben-Zeev, Fein, & Inzlicht, 2005; Spencer, Steele, & Quinn, 1999; Steele & Aronson, 1995; Vick, Seery, Blascovich, & Weisbuch, 2008; Woodcock, Hernandez, Estrada, & Schultz, 2012). Stereotype threat is the performance-diminishing apprehension of fulfilling an applicable, negative, ability-related stereotype in the face of a challenging/frustrating task (Aronson, Quinn, & Spencer, 1998). This means that commonly held prejudices in the United States, such as the view that Caucasian individuals are superior in overall intellect, can cause African Americans to underperform merely out of concern that they may underperform and thus confirm stereotypes. Stereotype threat therefore demonstrates the power that negative stereotypes wield over their targets.

In the undergraduate populations in which it has already been documented, stereotype threat shows a particular pattern of findings. Using the general example of verbal skills in African American and Caucasian undergraduates (based on Steele & Aronson's seminal 1995 study), stereotype threat produces the following interaction: When told that a test modeled after the Verbal Graduate Records Exam ([GRE®]; Educational Testing Service, 2010) measures "verbal ability" (stereotype threat [ST] condition), African Americans perform significantly lower than Caucasians on the test. On the other hand, when told that the same test measures "verbal problem-solving" (reduced-threat [RT] condition), the performances of the African Americans and Caucasians do not differ significantly. Nussbaum and Steele (2007) extended this manipulation to "academic ability" more broadly, characterizing a 20-question anagram measure as "a test designed to be diagnostic of academic ability" (ST condition) or "not a diagnostic test; an activity we use as a warm-up for problem-solving exercises" (RT condition; p. 129). Similar ST manipulations have been used to document stereotype threat in Latino (Woodcock et al., 2012) and low-socioeconomic-status (SES) undergraduates (Croizet & Claire, 1998; Spencer & Castano, 2007). A large meta-analysis has even demonstrated a "latent ability" effect whereby non-Asian ethnic minorities and women in quantitative fields participating in a "safe" (i.e., RT condition) significantly outperform controls in the RT condition (Walton & Spencer, 2009).

Studies of women's undergraduate mathematical performance have also consistently demonstrated the ST effect (i.e., that women perform significantly more poorly than men in the ST condition but the same as, or better than men in the RT condition) using manipulations similar to those used to examine stereotype threat

the already mentioned populations (Ben-Zeev et al., 2005; Carr & Steele, 2009; Good, Aronson, & Harder, 2008; Shih, Pittinsky, & Ambady, 1999; Spencer et al., 1999; Vick et al., 2008). For instance, Carr and Steele (2009) demonstrated stereotype threat in women when a math test was characterized as diagnostic of math/spatial ability and participants had to indicate their gender before taking the test (ST condition) relative to the RT condition, when the test was characterized as a "puzzle-solving exercise" and participants reported their gender after the test. Danaher and Crandall (2008) found that the manipulation of varying the order of asking participants to report gender information in and of itself was sufficient to replicate the ST effect (i.e., reporting gender before was the ST condition; reporting after was the RT condition). In addition, Vick et al. (2008) employed the manipulation for women's mathematical performance that the test about to be taken had "shown" (ST condition) or "not shown" (RT condition) "gender differences in performance in previous studies" (p. 627).

Even Caucasian males are not immune to stereotype threat. Aronson et al. (1999) demonstrated this by invoking a comparison of Caucasian males' mathematical performance to that of Asians. Similarly, Beasley and Fischer (2012) found the ST effect in white males majoring in science, technology, engineering, or math by asking questions to raise their race consciousness (e.g., "If I don't do well people will look down on others like me," p. 436). Moreover, a number of studies document stereotype threat in white males in non-academic domains. For a review of these studies and the ST literature overall, see Kit, Tuokko, and Mateer (2008).

One significant attribute of stereotype threat is that, in some contexts, it does not have to be explicitly invoked. In other words, if a stereotype-targeted population is told nothing but the standard instructions for a challenging stereotype-relevant test, they perform on the test as if stereotype-threatened (Osborne, 2001; Quinn & Spencer, 2001, Study 1; Spencer et al., 1999). For instance, Quinn and Spencer (2001) found that women performed significantly worse than men on an assessment of mathematical problem-solving when there was no mention of stereotypes or gender; however, there was no gender difference in an RT condition in which participants were told that the test "does not find gender differences." This finding means that negatively stereotyped students may be experiencing stereotype threat in routine stereotype-relevant academic situations, not just in assessment settings characterized by the marked demand characteristics of the experimental literature. Moreover, studies such as Good et al. (2008) and Walton and Spencer (2009), which find the performance of negatively stereotyped groups to exceed that of groups not negatively stereotyped when in a "safe" testing condition, present the prospect that educators may have set an artificially low ceiling for students vulnerable to stereotype threat in expecting performance only equal to that of non-threatened individuals.

In contrast to the steady stream of studies documenting the ST effect itself, fewer studies have examined the possible mediating variables in stereotype-threatened performance, and these studies have presented somewhat weak and conflicting findings (Kit et al., 2008). Potential mediators and moderators of the ST effect that have received experimental scrutiny include test anxiety, evaluation apprehension, cognitive interference, self-doubt, working memory, arousal, emotional regulation, and perserveration (Kit et al., 2008; Smith, 2004). Physiological measures include blood pressure (Blascovich, Spencer, Quinn, & Steele, 2001; Osborne, 2007) and constriction or dilation of the vasculature (Vick et al., 2008). Moreover, in the case of women taking math tests, neural networks associated with social and emotional processing had more heightened activation in an ST condition relative to women in an RT condition, whose networks associated with math learning were more activated (Krendl, Richeson, Kelley, & Heatherton, 2008).

A schematic representation of the theory behind the ST effect is provided in Figure 1. Included here is an indication of the prerequisite conditions for the operation of the effect (i.e., participant and test conditions), the various hypothesized mediating variables, and the outcomes of the effect. In contrast to self-fulfilling prophecy, which results in the individual's reduction in effort (Merton, 1948, as cited in Madon, Jussim, & Eccles, 1997), a stereotype-threatened individual is strongly identified with the academic area being evaluated and believes that he or she can be successful, thus working very hard to disprove or "rise above" the relevant, negative stereotype that applies to him or her.

Stereotype Threat and Learning Disabilities

Although the above review of research on stereotype threat is suggestive of the possibility that students

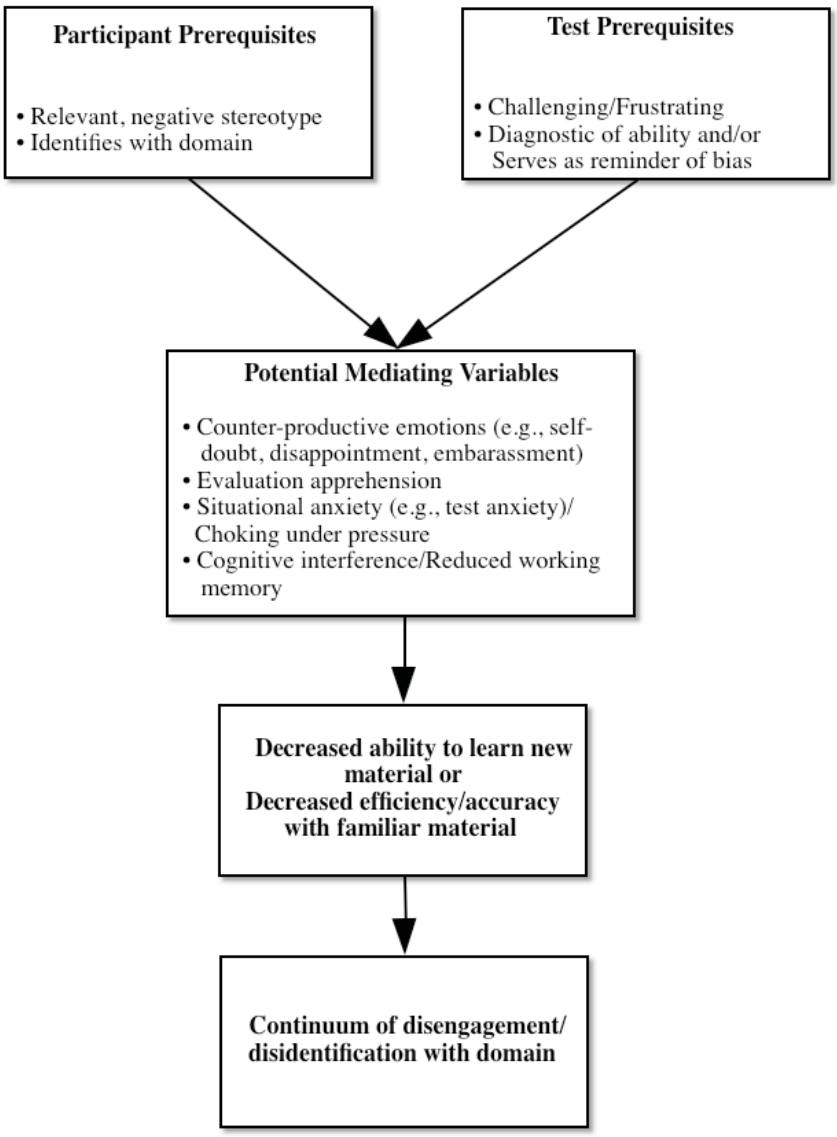

Figure 1. A Working Model of Stereotype Threat

with LD may be prime targets for stereotype threat, no studies have focused specifically on this population. However, like the populations targeted to date in the ST research, students with LD are the object of common stereotypes regarding academic ability, and they have been shown to differ from their peers on several of the factors hypothesized to mediate the ST effect. For example, numerous studies document the key role played by higher levels of test anxiety in these students than in those without LD, especially at the college level (Davis et al., 2009; Holzer et al., 2009;). In addition, memory limitations are a common processing deficit found in students with LD (Swanson & Saez, 2003), and recent studies have identified working memory as being compromised by stereotype threat (Schmader & Johns, 2003).

A few of the variables that have been investigated in previous ST studies seem particularly relevant to undergraduates with LD. One such factor is cognitive efficiency. In particular, Steele and Aronson (1995) found that African Americans who were administered a test that was purported to be diagnostic of verbal ability spent significantly more time per item than African Americans who were administered one purported to be non-diagnostic of ability, and more time than Cauca-

sians in both conditions. Even though all participants were scored only on those items they completed, stereotype threat had significantly decreased the accuracy of the African American participants as well, meaning that inefficiency due to ST, and not inadequate time per se, had been the source of these students' difficulties. It is noteworthy that a common accommodation for students with LD is extra time, based on the assumption that many of these students have significant deficits in speed of processing. It may be the case, therefore, that stereotype threat, in addition to an intrinsic processing inefficiency, contributes to slower test performance in undergraduates with LD.

Moreover, undergraduates with LD meet the two participant characteristics for susceptibility to stereotype threat elucidated by Aronson et al. (1998). First, they are subject to a number of negative stereotypes that are especially applicable in testing situations, such as that they have lower mental ability (May & Stone, 2010; Shapiro & Margolis, 1988). It is therefore possible that slight modifications of the verbal-ability/verbal-problem-solving manipulation would produce a similar ST effect in undergraduates with LD. Second, most studies of stereotype threat underscore the importance of participants being "identified with" a particular domain in order to be stereotype threatened (Osborne, 2001). That is, they must not only perceive themselves as having skills in a particular domain, but also view those skills as constituting part of their identities (Steele, 1999). Individuals with LD meet this second characteristic as well: Despite years of academic failure, they do not discount the importance of academic performance (Cosden & McNamara, 1997; Elbaum & Vaughn, 2003). Third, some findings indicate that ST effects exist in the presence of invisible social differentiation as well as in the case of visible factors such as gender or race. For example, students with low SES have been shown to evidence ST effects (Croizet & Claire, 1998; Woodcock et al., 2008), as have white undergraduates at Princeton from less common high schools (Alter, Aronson, Darley, Rodriguez, & Ruble, 2010).

In addition, noting that, "Most stereotype threat literature has focused on visible stigmatized status" (p. 142), Kit et al. (2008) summarized ST effects documented in populations with a traumatic brain injury, mental illness, and drug use. They proposed, "Contextual and psychological factors (i.e., negative stereotypes), in addition to organic causes, may be influencing…test performance for neuropsychologically compromised individuals" (p. 141). These findings lend further credence to the possibility that students with LD, another group with invisible differences, may also be affected. Finally, a number of studies are finding that stereotype threat interferes not just with participants' performance on tasks but also with their *learning* of new skills (Appel, Kronberger, & Aronson, 2011; Inzlicht & Ben-Zeev, 2003; Mangels, Good, Whiteman, Maniscalco, & Dweck, 2012; Taylor & Walton, 2011). Building on DeDeppo's (2009) assertion of a need to integrate academic/cognitive and affective/contextual factors in explaining college success for students with LD, the time therefore seems overdue to investigate the possible role of stereotype threat in undergraduates with LD.

Contrary to the situation for most other populations studied within the ST paradigm, however, it is important to note that there are actually intrinsic (i.e., non-stereotype-related) reasons as well as stereotype-related reasons to expect diminished performance. By definition, students with LD exhibit various processing deficits that would be expected to interfere with performance. Thus, this study's hypothesis is that the performance of undergraduates with LD on a disability-relevant task will significantly improve when stereotype threat is lifted (in the RT condition). In contrast to studies of stereotype threat in previous populations, however, the improvement will not result in performance as good as that of the controls in the RT condition because extended time limits will not be provided to undergraduates with LD.

The Present Study

Despite the preceding indications of the plausible role of ST in the poor performance of participants with LD, this potential role has not been explored. As a result, the primary goal of the present study was to provide a preliminary examination of the role of stereotype threat in the performance of students with LD. Using the same experimental paradigm as most previous studies of ST, students with and without a history of LD classification were assigned randomly to an ST or RT condition for participation in an academic assessment, and their relative performance was assessed. Given the emphasis placed on cognitive efficiency as a possible mediating factor in ST, both accuracy and speed of performance were examined for possible ST effects. To address these issues, we posed the following research questions:

1. Is the ST effect present in undergraduates with LD in a situation tapping academic skills?
2. Is any ST effect evident in speed, accuracy, or both?

Method

Participants

The participants for the study included 29 postsecondary students with LD and 62 without LD (hereafter, NLD) ranging from 18 to 24 years (LD: $M = 20.55$; $SD = 1.64$; NLD: $M = 20.31$; $SD = 1.06$). The NLD group and group with LD did not differ significantly in age, $t(89) = .86$, $p = .39$. Undergraduate-level students were chosen because the participants in most previous ST studies have been undergraduates, and thus it seemed prudent to examine the issue in this age group initially. Participants were recruited from two large Midwestern public research-intensive universities (hereafter referred to as University A and B) comparable to those used in past ST research. A total of 33 participants (12 LD; 21 NLD) attended University A, while 58 participants (17 LD; 41 NLD) attended University B. The sample size of students with LD was based on the number of participants in most previous ST studies, which generally had the same number or fewer participants than the present one. We felt it appropriate to treat the data from the two universities as a single sample because University A and B were comparable in undergraduate population (A: 24,493 students; B: 29,301 students), mean high-school GPA (A: 3.7; B: 3.6), and mean English ACT® score (25th - 75th percentile range for A was 26 – 30; for B: 25 – 29; http://www.usnews.com/usnews/edu/college/coworks.htm). Moreover, in the present study, the two school samples were not found to differ on ACT® scores, $t(85) = 0.87$, $p = .39$, age, $t(89) = 01.08$, $p = .28$, or number correct on the Verbal GRE® task used in this study, $t(89) = 1.22$, $p = .23$.

Participants were Caucasian males and females. The decision to exclude ethnic minorities but not women as participants was informed by two facts. First, ST has been documented in non-Asian minority students using portions of the Verbal GRE® and other "verbal-ability" tests (Blascovich et al., 2001; Nussbaum & Steele, 2007; Taylor & Walton, 2011). It has also been demonstrated in Asian women on a portion of the Quantitative GRE® (Shih et al., 1999). As a result, including ethnic minorities risked confounding race-related ST effects with LD-related ST effects. Second, although stereotype threat has been found to affect women taking math tests (Ben-Zeev et al., 2005; Carr & Steele, 2009; Krendl et. al., 2008; Mangels et al., 2012), gender effects have not been reported thus far in studies examining verbal skills in Caucasian, non-Asian minority, or low-SES students (Spencer & Castano, 2007; Taylor & Walton, 2011).

University A's policies and procedures related to students with disabilities list three criteria that must be met in order to be documented as having an LD. A student with LD must (a) come forward with a concern about his/her academic performance, (b) exhibit academic achievement that is significantly below expectation, and (c) demonstrate one or more areas of achievement at least one standard deviation discrepant from his/her overall or Verbal IQ. In addition, the student's assessment must have been conducted within the past three years.

University B's disability resource center cites similar criteria. Students seeking services must document a functional limitation that "significantly interfere[s] with…current academic performance." A significant functional limitation is "usually defined as a discrepancy of more than 2 standard deviations between achievement and IQ." Like University A, a student's documentation must also provide justification for requested accommodations and be no older than three years.[1]

At the end of the experiment all participants were asked to indicate which, if any, type of LD they had (see Measures). It was deemed essential to delay this question until the end of the experiment, as explicitly priming students' thoughts about their LD before the Verbal GRE® task would likely have interfered with the ST manipulation (Carr & Steele, 2009; Danaher & Crandall, 2008). Participants reported a variety of LD. The most common weakness was in the area of reading (21 students), followed by writing (12), language (9), math (4), attention (5), nonverbal (2), and other (3). Most students (17/29) reported their LD to affect mul-

1 To replicate the previous ST literature, it was necessary to identify students with LD at highly competitive universities. Students who were admitted to these schools and who submitted documentation of learning disabilities tend to match a gifted and LD profile (via the discrepancy model) in which achievement is just below average while intelligence is well above average. However, since what is at issue here is the state of being perceived (by self and others) as belonging to the LD target group, the specific nature of the individual's profile is not at issue.

tiple areas, with 1 student reporting 4 affected areas, 9 students reporting 3 areas, and 6 students reporting 2 affected areas.

Despite the number of participants reporting weaknesses in reading, students with LD who have made it to the level of competitive schools such as those used in this study were thought to be reading at an adequate level to take the Verbal GRE® task. Indeed, students with LD at competitive institutions may even have fairly high reading scores; however, this would not protect students with LD from ST. Steele (1999) noted that it is awareness of the stereotype and the knowledge that it applies to a group that give stereotype threat its power over targets. Targets do not have to believe the stereotype to be true of themselves and, in fact, they often work very hard to prove that it is not.

Steele's observation provided our rationale for not limiting the sample of students with LD to those with reading-specific LD. Simply the self-knowledge of having any LD means that participants (even if their disability is nonverbal or math-specific in nature) are aware that a common stereotype about the population to which they belong is that they are poor readers and less intelligent. As a result, the sample of students with LD was not limited in terms of type or severity of disability.

Measures

Who-Are-You questionnaire. This questionnaire was developed to provide information regarding participants' level of identification with reading (the to-be-threatened domain). This information was important since high identification with the target domain is a theoretical prerequisite for the ST effect. Issues of confidentiality made it impossible in the present study to survey participants beforehand. Thus, based on studies such as Spencer et al. (1999), Inzlicht and Ben-Zeev (2003), and Carr and Steele (2009), participants responded to a Likert-style question (on a scale from 1 to 7): "It is important to me that I have good reading skills." Based on Spencer et al.'s suggestion, individuals with a rating between 5 to 7 on this question were designated as "highly identified," those with a rating of 3 to 5 were designated as "moderately identified," and those with a rating between 1 and 3 were designated as "weakly identified" with reading.

Because of concern that the reading-identification question might threaten students with LD, we embedded it in a set of non-academic, trait-related questions (e.g., "It is important to me that I have strong leadership skills"). These nonacademic questions were not analyzed. The Who-Are-You questionnaire was the first measure to which participants responded, and it was purposefully treated as separate from the main experiment. For instance, most participants filled this out while waiting for others in their testing session to arrive.

Demographic questionnaire. This questionnaire (see May & Stone, 2010, Appendix A) was created in order to determine the LD status of participants, as well as to determine other possible covariates. For the present study, the covariate deemed most important was English ACT® score (ACT, Inc., 2010). A number of ST studies (Spencer & Castano, 2007; Taylor & Walton, 2011) have used the Verbal SAT® (College Board, 2010) score as a covariate to control for individual differences in verbal ability.

English ACT®. Consistent with previous studies of stereotype threat, we collected each participant's self-reported score on the English ACT® or Verbal SAT® to be used as a possible covariate in the main ST analyses. (Since more participants reported taking the ACT® than the SAT®, Verbal SAT® scores were converted to English ACT® scores using tables for ACT®-SAT® concordance, 2010.)

Verbal GRE® task. A researcher-created task consisting of sample and discontinued items from the Verbal GRE® was used as the basis for examining effects of the ST manipulation on students' test performance. In addition, the time spent on each of these questions, which was recorded automatically by the software used for administration (see Procedure), was assessed, consistent with many previous ST studies. Many of these studies (Alter et al., 2010; Good et al., 2008; Osborne, 2007; Vick et al., 2008;) used portions of the Verbal or Math GRE® because they deemed these tests to be appropriately challenging to activate stereotype threat in undergraduates. Based on the predominance of reading difficulties in students with LD, as well as an informal survey of several undergraduates about their perceptions of individuals with LD, it was deemed appropriate to use the Verbal GRE® in the present study because more reading-related than math-related stereotypes seemed to exist about students with LD.

The 45-item mock Verbal GRE® contained four types of items: antonyms, reading comprehension, analogies, and sentence-completion. The criteria gov-

erning the proportion of each item type, the standard instructions for the Verbal GRE® sections, and the majority of the items (38) were taken from *GRE®: Practicing to Take the General Test, 9th Edition* (Educational Testing Services, 1998). The remaining seven items were taken from *Cliff's GRE® Preparation Guide* (Bobrow, Orton, & Covino, 1995). Participants were allowed 30 minutes to complete the test. This time limit was deemed appropriate based on previous ST research utilizing the GRE®, which employed time limits between 15 and 30 minutes.

Test items were selected from those answered correctly by 30-60% of students, based on information in the Educational Testing Services (1998) guide. We based this criterion on reports in the ST literature that items were answered correctly by 30-50% of participants. Finally, test length was based on pilot testing, which had revealed that this number of items would allow about half of the participants to finish. (The actual completion rates in the present study were 48% [NLD] and 33% [LD].)

Items were presented one at a time. Five response choices accompanied each item, along with a sixth choice that gave participants the option to "Skip Question." Each section of items was preceded by the standard instructions given by Educational Testing Services (1998). Participants read and were told that their score would be based on only those items they answered, and that they would receive no credit or penalty for items left blank. This information was included for two reasons: First, it represents the standard terminology of the GRE® instructions, and second, emphasizing the scoring criterion for items left blank was intended to alleviate concerns about not finishing that students with LD, in particular, might have had. In addition, Inzlicht and Ben-Zeev (2003) refer to this method of scoring as "consistent with the stereotype threat literature" (p. 800). See the Appendix for the GRE® instructions in the ST and RT conditions in the present study.

Manipulation check. Two 6-point Likert-scale questions served as a check of participants' perceptions of the condition to which they were assigned (i.e., a test of ability versus problem-solving). Analyses involving these items allowed us to verify that the results reported below were not an artifact of differential "buy-in" on the part of participants in the various groups. The instructions and questions for the manipulation check were modeled after Steele and Aronson (1995).

Procedure

Participant recruitment and selection. All participants were solicited by email with a message offering only a general explanation of the study's focus. The message indicated that the researcher was "a doctoral student at [another university] studying psychology, and asked if they "would be willing to participate in an experiment… [that] would require approximately an hour of [their] time, for which [they] would be paid $15/20." Participants were further told that they would "take a test and fill out a few questionnaires," and that, "[their] identity would in no way be connected to either [their] performance on the test or [their] questionnaire responses."

We solicited a total of 450 students with LD and 600 NLD students at University A. At University B, we solicited 227 students with LD and 450 NLD students. We solicited NLD students from the rosters of introductory educational psychology courses at each institution. In the case of students with LD, the solicitation came from disability service specialists. A total of 62 NLD students and 29 students with LD participated in the study.

It should be noted that neither the age, $t(89) = .86$, $p = .39$, nor the English ACT® scores, $t(85) = 1.49$, $p = .14$, of participants with and without LD differed in the present study. See Table 1 for these and other participant characteristics. Although the absence of differences on the ACT® may raise questions as to the existence of LD-NLD between-group differences, this absence of differences is likely attributable to the fact that almost all of the students with LD reported having been granted extra time on the ACT®. Eligibility for time accommodations on the ACT further substantiates the history of LD diagnosis for these students. Consistent with the ACT®'s policy for documentation, students in the LD group would have been required to submit documentation of high-school-approved accommodations to the ACT®. Moreover, it is important to note that the LD group performed significantly worse than the NLD group on the Verbal GRE® measure administered in the present study, for which no accommodations were provided (see Results). The effect size for mean difference in GRE® performance falls in the moderate (d = .55) to large (raw number correct, d = .70) range.

Assignment to condition. Participants were scheduled for testing in small groups in accordance with their availability. Each small group was assigned randomly to either the ST or RT condition. Since the

Table 1

Comparison of Participants with LD and Controls on Background Characteristics

	Age			Gender		English ACT® Score		
Participants	Range	Mean	SD	M	F	Range	Mean	SD
LD	18-24	20.55	1.64	7	22	17-35	25.54	3.95
Control	18-23	20.31	1.06	16	46	20-35	26.74	3.28

LD status of participants was not known until after testing was completed, it was not possible to control the exact proportions of LD and NLD students in each condition. At the completion of testing, a total of 17 students with LD and 32 NLD students had participated in the ST condition, while 12 students with LD and 30 NLD students had received the RT condition.

Task administration. At both campuses, students were asked to report to a computer lab/room in the School of Education in groups of between 3 and 11 for a single session lasting approximately an hour. The Who-Are-You questionnaire, as well as the Verbal GRE® task, were all presented on a Macintosh G4 computer using SuperLab Pro™ software (Cedrus Corporation, 1998). The ST manipulation was incorporated into the directions for the Verbal GRE®, which students read silently while the experimenter read aloud. The ST manipulation of "verbal ability" versus "verbal problem-solving" (often used with African Americans) was altered to a contrast between "verbal reasoning and reading abilities" versus "stylistic variations in problem-solving." This is because pilot testing and a companion study (May & Stone, 2010) found students with LD to be negatively stereotyped regarding their intellectual abilities and reading skills.

Participants were given 30 minutes to complete as many of the 45 GRE® items as they could. They then responded to the manipulation check and the demographic questionnaire. Because participants finished at different times, they were debriefed in writing with encouragement to ask questions, voice concerns, and share any suspicion they might have had about what the study was testing. They were paid $20 (at University A) or $15 (at University B) for their participation and then dismissed.

Cross-check of disability status. After the study was completed at each site, the names of all the students who participated in the study were sent to the disabilities services office so that they could determine the number of students with LD who had participated in the study. This served as a cross-check of the number who self-identified as having a learning disability. To maintain participant confidentiality and because both disability service offices had inadequate staff resources, no names or evaluation information for students were requested, only the number of students with LD appearing on the participant list. At both schools, the number sent by the disabilities specialists matched the number of students who had self-identified at the end of the study.

Scoring Data in Preparation for Analysis

Who-Are-You questionnaire. Based on the identification-with-reading question, 90% of the students with LD would be rated as highly identified, compared to 98% of their NLD peers. This suggests a high level of identification with reading in both samples. In addition, it is notable that 89 of the 91 participants reported having received a "B" or higher in their most recently taken humanities-related course.

Verbal GRE® Accuracy. To assess accuracy of performance on the GRE® task, we scored the data in the manner used by Educational Testing Services (1998), Inzlicht and Ben-Zeev (2003), Quinn and Spencer (2001), and Spencer et al. (1999): + 1 point for correct responses, - 1/5 point for incorrect responses. This method of scoring will be referred to as simply, "ETS Score." Items left blank were therefore not counted as incorrect in this means of scoring, meaning that students with LD, who were less likely to finish

Table 2

GRE® Performance of Participants with and Without LD

	Participants with LD		
Scoring Method	All	ST Condition	RT Condition
ETS Score	8.14 (5.06)	7.95 (4.42)	8.42 (6.05)
Time per Item (all except RC)	27.95 (7.04)	29.22 (8.18)	26.16 (4.79)
	Control Participants		
Scoring Method	All	ST Condition	RT Condition
ETS Score	11.03 (4.96)	10.82 (5.70)	11.25 (4.12)
Time per Item (all except RC)	24.43 (4.43)	23.84 (5.21)	25.07 (3.39)

Note. Numbers in parentheses are standard deviations. The unit for all three time-per-item measures is seconds. ETS Score = number of correct responses – 1/5 times the number of incorrect responses. ST = stereotype-threat; RT = reduced-threat; RC = reading comprehension.

the GRE® task, would not be penalized beyond not having the opportunity to attempt as many items as students without LD.[2]

GRE® Time. Because time spent per item has been found to increase under ST conditions in some studies (Aronson et al., 1999; Carr & Steele, 2009; Steele & Aronson, 1995), it was deemed important to measure this variable, which was recorded by the SuperLab Pro™ (Cedrus Corporation, 1998) software. Since students with LD, in particular, were predicted to take a disproportionately long time on the reading comprehension items, the average time spent per item was calculated in three ways for each student: for all items, for the reading comprehension items only, and for all items except the reading comprehension items (i.e., antonyms, analogies, and sentence completion).

Data Analyses. To address the main research questions, performance on the GRE® task as a function of ST condition by undergraduate students with and without LD was analyzed using separate 2 x 2 analyses of variance for accuracy and time spent per item. In addition, in supplemental analyses, level of identification with reading, English ACT® score, meta-stereotype of learning disabilities, credibility of ST manipulation, entity vs. incremental view of intelligence, and size of testing group, were evaluated for their potential role as covariates.

Results[3]

ST Effect on the Verbal GRE® Task

Accuracy. In order to test for a ST effect on the accuracy of GRE® performance, a 2 x 2 ANOVA (participant status by ST manipulation) was performed using the ETS Score.[4] There was a main effect for

2 We also used two other approaches to scoring accuracy: total number of items correct and proportion correct for items attempted. Since the results were comparable, we report only the ETS Score method here.

3 It is important to note that preliminary analyses of responses to the manipulation check items (see Methods) did not reveal any significant differences between the two participant groups in the believability of the ST/RT instructions (ST condition: LD 78%; NLD 79%; RT condition: LD: 46% and NLD 40%). In addition, follow-up analyses of the main ST findings indicated that the results reported below remained the same when restricted to those participants in the two groups who reported believing their condition.

4 English ACT® scores and GRE® scores for the entire sample were significantly correlated, $r(87) = .33$, $p = .002$. Thus, the English ACT® score appeared to be an appropriate covariate for analyses involving participants' accuracy on the Verbal GRE®. As a result, ANCOVAs incorporating this variable were used in initial analyses of students' performance. However, the results of these

participant status, $F(1, 87) = 6.18$, $p = .015$, but the ST manipulation effect, $F(1, 87) < 1$, $p = .70$, and the interaction , $F(1, 87) < 1$, $p = .98$, were not significant. (See Table 2.)[5]

Time per item. Preliminary analyses of all three approaches to scoring the time-per-item data revealed that students with LD in the RT condition took less time per item than did their counterparts in the ST condition, whereas controls took the same amount of time per item in both conditions, or took *more* time per item in the RT condition. However, an examination of the means for the time-per-item data revealed a high degree of variability. Upon closer examination, it was noted that this high variability was due largely to the reading comprehension items. Since all three ways of scoring the data yielded similar patterns, the average time per item on all items except reading comprehension items was used for this analysis because of its lower variability. In addition, we further reduced variability by removing item response times that were greater than 2.5 standard deviations from the mean for a given participant.

As expected, a main effect was found for participant status, $F(1, 87) = 7.07$, $p = .01$, such that students with LD ($M = 27.95$ seconds/item) took more time per item than controls ($M = 24.43$ seconds/item). Also as predicted, there was a marginally significant ST-manipulation x participant-status interaction, $F(1, 87) = 3.11$, $p = .08$. Two follow-up *t*-tests were then conducted. The results of these tests indicated that students with LD ($M = 29.22$ seconds/item) spent more time per item than controls ($M = 23.84$) in the ST condition, $t(47) = -2.81$, $p = .007$, but that they worked at essentially the same rate as controls in the RT condition (LD: 26.16; control: 25.07), $t(40) = -0.83$, $p = .41$ (see Figure 2).

analyses were comparable to those found in analyses without the covariate. Because of increased power resulting from the inclusion of subjects with missing ACT scores as well as ease of presentation and interpretation, we report the results without a covariate. We considered five other measures for possible use as covariates: level of identification with reading, credibility of the ST manipulation, entity vs. incremental view of intelligence, meta-stereotype of learning disability, and size of testing group. None of these measures correlated significantly with the GRE® measures, and only credibility of ST manipulation is discussed further.

5 It should be noted that the other two approaches to assessing accuracy of performance also revealed no main effect for threat condition and no interaction.

Discussion

A central assumption motivating the present study was that ST would be found to hinder the test performance of undergraduates with LD. This assumption was based on the fit between the characteristics of this population and the current working model of stereotype threat (see Figure 1). To date, no other studies have examined the role of ST in undergraduates with LD. This study, although exploratory in nature, expands the limited amount of research on undergraduate students with LD and extends the very small body of research on stereotype threat in populations with invisible differences (e.g., individuals from low-SES or atypical educational backgrounds).

Presence of Stereotype Threat in Undergraduates with LD

Contrary to expectation, the present study found only marginal support for the hypothesis that ST is a contributor to the diminished performance of undergraduates with LD. As predicted, these students tended to spend more time per test item in the ST condition than in the RT condition. In addition, their time per item in the RT condition did not differ significantly from that of controls in either the RT *or* the ST condition. This pattern of results suggests that students with LD were able to work as efficiently as controls when the burden of ST was lifted. However, the interaction between student status and threat condition was only marginally significant (.08), and the time-per-item differences did not lead to differential accuracy scores for the two groups of participants as a function of the ST condition. Our explanation for this null effect relates to the nature of the GRE® testing procedure, which allows participants to skip items without penalty. Indeed, students with LD skipped more items overall. Given the generous time allotted for the test, this behavior did not impose a significant penalty on those participants. Thus, the overall performance differences in ETS Score between students with and without LD were attributable to differing numbers of items completed, rather than to differential success on items attempted.

Therefore, consistent with numerous studies regarding test performance by students with LD, this study found undergraduates with LD to take significantly longer than their non-LD counterparts on a test, regardless of the testing condition. Moreover, the present study provides suggestive evidence that ST (not

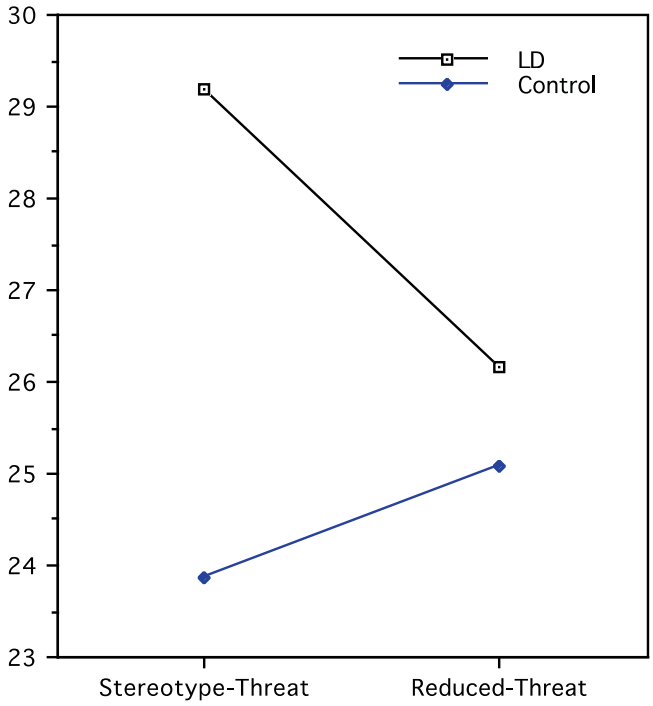

Figure 2. Main Effect and Interaction for Time per Item.

just the processing weaknesses intrinsic to students with LD) may contribute to the slower speed of test performance typically documented in undergraduates with LD. Clearly, more research is needed to validate this possibility, but the present results suggest that such research is a pressing need.

A Check for the Prerequisites of Stereotype Threat

As a further means of exploring the likely reasons for the weak findings in the present study regarding the presence of ST in undergraduates with LD, the components of the current ST model were revisited to ensure that all prerequisites had been met (see Figure 1).

One prerequisite for susceptibility to ST is that the individual is identified with the performance domain in question. This prerequisite appears to have been satisfied, in that no participants with LD were observed to report low identification with reading (as measured by their responses on the measure described above).

The second prerequisite for susceptibility to ST is that the test administered should be challenging, "push[ing] ability to the limit" (Aronson et al., 1998, p. 92). For all groups and conditions in the present study, subjects performed at between 40 to 42% accuracy. This figure is consistent with those reported in many studies reporting significant ST effects (Inzlicht & Ben-Zeev, 2003; Krendl et al., 2008; Shih et al., 1999; Steele & Aronson, 1995) although some report using easier tests (Alter et al., 2010) and others report using harder ones (Osborne, 2007). Nevertheless, the item difficulty and performance of participants in the present study seem appropriate based on the majority of ST literature.

The third prerequisite for susceptibility to ST is that the test administered should purport to measure ability. Based on an examination of responses to the manipulation-check items, the majority of participants did indeed tend to believe that the Verbal GRE® in the present study was a test of their ability in the ST condition. Re-running the analyses on those students who reported believing the ST and RT instructions produced the same findings as when these analyses were performed for all participants. Failure to satisfy the ability prerequisite therefore does not therefore seem to be a likely explanation for the negative findings.

The fourth prerequisite is that there exists a rel-

evant, negative stereotype about individuals with LD. Certainly negative stereotypes have been documented in some previous studies focused on individuals with LD (Houck, Asselin, Troutman, & Arrington, 1992; Shapiro & Margolis, 1988), as well as in a companion study by the authors (May & Stone, 2010) using a larger sample that included the participants in the present study. In that study, 112 of the 138 participants (about 81%) reported meta-stereotypes of individuals with LD that were negative (see May & Stone, 2010, for details).

It is noteworthy that the specific type of intellectual deficiency reported by participants in the companion study (May & Stone, 2010) was less often related explicitly to reading and/or verbal abilities, suggesting possible ambiguity regarding potential threat situations. Also, 47% of the individuals with LD reported a meta-stereotype of LD that did not relate explicitly to intelligence. Because an individual's assumptions regarding the stereotype that others hold about his/her group moderates his/her susceptibility to ST (Hamilton & Sherman, 1994), the heterogeneous, non-intelligence related meta-stereotypes of LD that some undergraduates with LD held may have weakened any influence of the ST manipulation.

Thus, only three of the four prerequisites for activating ST appear to have been fully satisfied in the present study: that all participants were adequately identified with reading LD, that the test administered was challenging, and that the test was believed to be an assessment of ability. The remaining prerequisite, that a negative, ability-related stereotype about students with LD existed in the present sample, appears to have been only partially met. Therefore, the findings related to this prerequisite deserve future exploration as an alternate explanation for the pattern of findings in the present study.

Alternate Explanations and Implications of the Findings

In the present section, two possible explanations for the failure to find the typical performance decrements associated with ST are discussed: limitations in statistical power, and the representativeness of the sample with LD who participated in the present study.

Power Limitations. One alternative explanation for the mixed findings regarding ST is that the GRE® task used in the present study was not sufficiently sensitive (at least not in the 30-minute time limit employed here) to detect the subtle performance differences between the ST and RT conditions that were originally predicted. In addition, there may have been an insufficient number of participants (especially those with LD) to document this effect. With regard to this first concern, it is noteworthy that most previous researchers who administered graduate-level tests (see Introduction) instituted time limits ranging from 15 to 30 minutes.

In order to evaluate the speculation that there may have been an insufficient number of participants with LD in the present study, it was necessary to calculate effect sizes for previous ST studies and to apply these to the present study. For instance, using Cohen's (1988) power tables and estimating the means and standard deviations using a graph in Steele and Aronson (1995), it was determined that Steele and Aronson's effect size for detecting the simple main effect (of African Americans in the RT condition performing significantly better than in the ST condition) was .54. Data from Spencer et al.'s (1999) study of women's mathematical performance yielded fairly comparable effect sizes (i.e., a main effect size of .65 and an interaction effect size of .47). Using the more conservative effect sizes from Steele and Aronson's (1995) study to calculate the present study's power to detect the simple main effect (that students with LD would perform significantly better in the RT than the ST condition) yielded a power of .99. Thus, the power to demonstrate the hypothesized effects in the present design appears high.

The representativeness of the LD sample. A second possible explanation for this study's failure to find the ST effect for students with LD may relate to the participant recruitment process. Because of Institutional Review Board guidelines, the solicitation email sent to students made specific reference to the fact that students would take a "short test." It is therefore possible that those students with LD who were most likely to feel threatened by the prospect of taking a test would not respond to the participant-solicitation email in the first place. Additional evidence that the group of participants with LD in the present study may have been less threatened by the prospect of taking a test is that these students did not differ significantly from controls in their level of state test anxiety (unpublished data). This finding is in stark contrast to those in a number of other studies of test anxiety, in which undergraduates with LD demonstrated significantly higher levels of test anxiety (Holzer et al., 2009; Hoy et al., 1997).

It should also be noted that the reliance on a discrepancy model of LD by the institutions from whom

we drew participants is another potential source of bias in the study. As noted by Sparks and Lovett (2009b), discrepancy models are widely used in the eligibility decisions of postsecondary institutions. The data reported by Sparks and Lovett indicate that the individuals identified by such a method overlap only partially with those identified by other methods such as low achievement. As noted in the earlier discussion of participant selection (see Methods), the ST framework assumes that individuals who identify with a threatened population are susceptible to the relevant ST regardless of their actual characteristics. Thus, the participants' achievement status should not be a deciding factor in their susceptibility to ST if indeed the phenomenon exists in the LD population. Nonetheless, the existence of stereotype threat should be explored in individuals who fit the classic low-achievement pattern as well as in high-functioning individuals with LD such as those in the current sample.

Directions for Future Research

The findings from this study represent an important starting point in the analysis of the role of ST in the LD population; however, they need to be replicated and expanded. In particular, we suggest several design issues to consider in future studies.

One primary limitation of the present study is the low number of participants with LD as well as the inherent self-selection bias involved in limiting the study to students with LD who have registered with their institution's disability services office. Addressing this issue in future studies is important; however, there are challenges in doing so. College students with LD represent a small, albeit growing population (Orr & Hammig, 2009; Sparks & Lovett, 2009a). In addition, it would be expected that, if ST does indeed impact students with LD, the potential study population would be limited further by the fact that few such students would want to participate in a study that would arouse the degree of discomfort typically associated with ST. One possible approach to addressing this issue would involve accessing potential participants via the graduation records of high school special education departments.

Two additional possibilities for modifying the ST manipulation are to use an identity-priming manipulation or a test-bias manipulation instead of the task-reframing manipulation used in the present study. For instance, Vick et al. (2008) used the ST manipulation of telling female participants in their study that the challenging math test they were about to take had shown gender differences in performance in previous studies. Such a manipulation could easily be reworded to read, "This test has previously shown/not shown performance differences between students with and without LD." Along similar lines, an ST manipulation similar to the one used by Danaher and Crandall (2008) could be modified and used with participants with LD. Specifically, they demonstrated the ST effect by simply asking or not asking female participants to self-identify their gender before taking the Quantitative GRE®, and this could easily be modified by asking students to self-identify their LD status before taking the Verbal GRE®.

Implications for Practice

This study extends the current body of research supporting the accommodation of providing extra time to students with LD (Alster, 1997; Lindstrom & Gregg, 2007). Specifically, our study found that students with LD tended to spend significantly more time per item than controls in the ST condition, and the reader will recall that the typical test administered at the college level is assumed by default to be one of ability (and therefore akin to an ST condition) by the students who take it (see earlier discussion of work by Spencer et al., 1999). Because ST is by default in the "on" position in undergraduate tests and because many students, even those with invisible differences, appear vulnerable to negative stereotypes regarding their performance, research on the test performance of all students should be undertaken to determine how factors such as ST may cause educators to underestimate the academic performance of many.

Conclusions

Based on the results of this study, the role of ST in the test performance of college students with LD remains uncertain. Although there was no decrement in the overall number of items answered correctly, it is nonetheless noteworthy that students with LD tended to spend more time per item and to skip more items. This finding suggests that reducing the amount of ST in a testing situation for students with LD might result in performance that is no less accurate but more efficient than typical testing situations and might lead to less emotional discomfort and/or distraction.

References

ACT®, Inc. (2010). *The ACT® Assessment.* Iowa City, IA: Author.

ACT®-SAT® concordance. (2010). Retrieved September 9, 2010 from the ACT, Inc. website: http://www.act.org/aap/concordance/

Appel, M., Kronberger, N., & Aronson, J. (2011). Stereotype threat impairs ability building: Effects on test preparation among women in science and technology. *European Journal of Social Psychology, 41*, 904-913.

Alster, E. H. (1997). The effects of extended time on algebra test scores for college students with and without learning disabilities. *Journal of Learning Disabilities, 30*, 222-227.

Alter, A. L., Aronson, J., Darley, J. M., Rodriguez, C., & Ruble, D. N. (2010). Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. *Journal of Experimental Psychology, 46*, 166-171.

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental and Social Psychology, 35*, 29-46.

Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. Swim & C. Stangor (Eds.), Prejudice: *The target's perspective* (pp. 83–103). Boston: Academic Press.

Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math, and engineering majors. *Social Psychology of Education, 15*, 427-448.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Psychology, 41*, 174-181.

Blascovich, J., Spencer, S .J., Quinn, D., & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science, 12*, 225-229.

Bobrow, J., Orton, P. Z.,& Covino, W. A. (1995). Cliffs Graduate Record Examination General Test Preparation Guide. Lincoln, NB: Cliffs Notes, Inc.

Carr, P., & Steele, C.M. (2009). Stereotype threat an inflexible perseverance in problem solving. *Journal of Experimental Social Psychology, 45*, 853-859.

Cedrus Corporation (1998). *SuperLab Pro: Experimental laboratory software* (Version 1.74) [Computer software]. Phoenix, AZ: Author.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

College Board (2010). *The SAT®*. New York: Author.

Cosden, M. A., & McNamara, J. (1997). Self-concept and perceived social support among college students with and without learning disabilities. *Learning Disability Quarterly, 20*, 2-12.

Croizet, J.-C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*, 588-594.

DaDeppo, L. M. W. (2009). Integration factors related to the academic success and intent to persist of college students with learning disabilities. *Learning Disabilities Research & Practice, 24*, 122-131.

Danaher, K., & Crandall, C. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*, 1639-1655.

Davis, T. E., Nida, R. E., Zlomke, K. R., & Nebel-Schwalm, M. S. (2009). Health-related quality of life in college undergraduates with learning disabilities: The meditational roles of anxiety and sadness. *Journal of Psychopathology and Behavioral Assessment, 31*, 228-234.

DeSimone, J .R., & Parmar, R. S. (2006). Middle school mathematics teachers' beliefs about inclusion of students with learning disabilities. *Learning Disabilities Research & Practice, 21*, 98-110.

Educational Testing Service (1998). *GRE®: Practicing to take the general test* (9th ed.). Princeton, NJ: Educational Testing Service.

Educational Testing Service (2010). *GRE® General Test.* Princeton, NJ: Author.

Elbaum, B., & Vaughn, S. (2003). For which students with learning disabilities are self-concept interventions effective? *Journal of Learning Disabilities, 36*, 101-108.

Good, C., Aronson, J., & Harder, J.A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*, 17-28.

Hartman-Hall, H. M., & Haaga, D. A. F. (2002). College students' willingness to seek help for their learning disabilities. *Learning Disability Quarterly, 25*, 263-274.

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In T. K. Krull & R. S. Wyer (Eds.), *Handbook of social cognition* (2nd ed., Vol. 2, pp. 1-68). Hillsdale, NJ: Lawrence Erlbaum.

Holzer, M. L., Madaus, J. W., Bray, M. A., & Kehle, T. J. (2009). The test-taking strategy intervention for college students with learning disabilities. *Learning Disabilities Research & Practice, 24*, 44-56.

Houck, C. K., Asselin, S. B., Troutman, G. C., & Arrington, J. B. (1992). Students with learning disabilities in the university environment: A study of faculty and student perceptions. *Journal of Learning Disabilities, 25*, 678-684.

Hoy, C., Gregg, N., Wisenbaker, J., Manglitz, E., King, M., & Moreland, C. (1997). Depression and anxiety in two groups of adults with learning disabilities. *Learning Disability Quarterly, 20*, 280-291.

Inzlicht, M., & Ben-Zeev, T. (2003). Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology, 95*, 796-805.

Kit, K. A., Tuokko, H. A., & Mateer, C. A. (2008). A review of the stereotype threat literature and its application in a neurological population. *Neuropsychological Rev, 18*, 132-148.

Krendl, A. C., Richeson, J. A., Kelley, W. M., & Heatherton, T. F. (2008). The negative consequences of threat: A functional magnetic resonance imaging investigation of neural mechanisms underlying women's underperformance in math. *Psychological Science, 19*, 168-175.

Lindstrom, J. H., & Gregg, N. (2007). The role of extended time on the SAT® for students with learning disabilities and/or attention-deficit/ hyperactivity disorder. *Learning Disabilities Research & Practice, 22*, 85-95.

Madon, S., Jussim, L., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology, 72*, 791-809.

Mangels, J. A., Good, C., Whiteman, R. C., Maniscalco, B., & Dweck, C. S. (2012). Emotion blocks the path to learning under stereotype threat. *SCAN, 7*, 230-241.

May, A. L., & Stone, C. A. (2010). Stereotypes of individuals with learning disabilities: Views of college students with and without learning disabilities. *Journal of Learning Disabilities, 43*, 483-499.

Murray, C., Goldstein, D. E., Nourse, S., & Edgar, E. (2000). The postsecondary school attendance and completion rates of high school graduates with learning disabilities. *Learning Disability Research and Practice, 15*, 119-127.

Murray, C. Wren, C. T., & Keys, C. (2008). University faculty perceptions of students with learning disabilities: Correlates and group differences. *Learning Disability Quarterly, 31*, 95-113.

Nussbaum, A. D., & Steele, C. M. (2007). Situational disengagement and persistence in the face of adversity. *Journal of Experimental Social Psychology, 43*, 127-134.

Orr, A. C., & Hammig, S. B. (2009). Inclusive postsecondary strategies for teaching students with learning disabilities: A review of the literature. *Learning Disability Quarterly, 32*, 181-196.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology, 26*, 291-310.

Osborne, J. W. (2007). Linking stereotype threat to anxiety. *Educational Psychology, 27*, 135-154.

Quinlan, M. M., Bates, B. R., & Angell, M. E. (2012). What can I do to help? Postsecondary students with learning disabilities' perceptions of instructors' classroom accommodations. *Journal of Research in Special Educational Needs, 12*, 224-233.

Quinn, D. M., & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues, 57*, 55-71.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85*, 44-452.

Shapiro, A., & Margolis, H. (1988). Changing negative peer attitudes toward students with learning disabilities. *Reading, Writing, and Learning Disabilities, 4,* 133-146.

Shih, M., Pittinsky, T., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.

Smith, J. L. (2004). Understanding the process of stereotype threat: A review of meditational variables and new performance goal directions. *Educational Psychology Review, 16*, 177-206.

Sparks, R. L., & Lovett, B. J. (2009a). College students with learning disability diagnoses: Who are they and how do they perform? *Journal of Learning Disabilities, 42*, 494-510.

Sparks, R. L., & Lovett, B. J. (2009b). Objective criteria for classification of postsecondary students as learning disabled. *Journal of Learning Disabilities, 42,* 230-239.

Spencer, B., & Castano, E. (2007). Social class is dead. Long live social class! Stereotype threat among low socioeconomic status individuals. *Social Justice Research, 20*, 418-432.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental and Social Psychology, 35*, 4-28.

Steele, C. M. (1999, August). Thin ice: 'Stereotype threat' and black college students. *The Atlantic Monthly, 284*, 44-54.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.

Swanson, H. L., & Saez, L. (2003). Memory difficulties in children and adults with learning disabilities . In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of Learning Disabilities* pp. 182-198). New York: Guilford.

Taylor, V. J., & Walton, G. M. (2011). Stereotype threat undermines academic learning. *Personality and Social Psychology Bulletin, 37*, 1055-1067.

Vick, S. B., Seery, M. D., Blascovich, J., & Weisbuch, M. (2008). The effect of gender stereotype activation o n challenge and threat motivational states. *Journal of Experimental Social Psychology, 44*, 624-630.

Vorauer, J. D., Main, K. J., and O'Connell, G. B. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *Journal of Personality and Social Psychology, 75*, 917-937.

Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science, 20*, 1132-1139.

Woodcock, A., Hernandez, P. R., Estrada, M., & Schultz, W. (2012). The consequences of chronic stereotype threat: Domain disidentification and abandonment. *Journal of Personality and Social Psychology, 103*, 635-646.

Zurcher, R., & Bryant, D. P. (2001). The validity and comparability of entrance examination scores after accommodations are made for students with LD. *Journal of Learning Disabilities, 34*, 462-471.

About the Authors

Alison May received her Ph.D. in Communication Sciences and Disorders-Learning Disabilities from Northwestern University in 2002, where she had the privilege of being mentored by her co-author (then dissertation chair). Her professional experience includes directing Student Support Services TRiO programs at Columbia College Chicago and Oakton Community College between 2001-2007. She also served as adjunct faculty in the Joan and Irving Harris Center for Early Childhood Education at Columbia College from 2003-2009. In addition, she is a certified special education teacher (Learning & Behavioral Specialist). Alison currently serves as Assistant Dean of Students and Director for the Office of Services for Students with Disabilities at Northwestern University, where she has worked since 2007. Her research interests focus on the psychosocial context of academic achievement and well-being in at-risk populations, especially undergraduate students with disabilities. She can be reached by email at: a-may@northwestern.edu.

Addison Stone earned his Ph.D. in Developmental Psychology from the University of Chicago in 1976. From 1978-2000, he served as a faculty member in the Department of Communication Sciences and Disorders at Northwestern University, where he supervised clinical practica in the differential diagnosis of learning problems and directed the graduate programs in Learning Disabilities. From 2000-2012, he was Professor and Chair of Educational Studies at the University of Michigan. His recent research has focused on social contextual influences on the development of children with language and learning disabilities. He has served as co-editor of *Learning Disabilities Research & Practice*, and *Elementary School Journal*, and as a consulting editor with Guilford Press for an interdisciplinary series on Challenges in Language and Literacy Development. He can be reached by email at: addisons@umich.edu.

Authors' Note

This study was made possible with the generous support of a Donald D. Hammill Foundation Research Scholarship, a Northwestern University Dissertation-Year Fellowship, and a Northwestern University Graduate Research Grant.

Appendix

Key Stereotype-Threat (ST) and Reduced-Threat (RT) Instructions from the Verbal GRE® Task Procedure

(Note: The wording differences between the two conditions are bracketed and underlined, and instructions were read aloud by the experimenter while participants read them silently.)

This study is concerned with [ST: various personal factors involved in performance on problems requiring verbal reasoning and reading abilities] [RT: stylistic variations in problem-solving]. The items on the test you are about to take are similar in format to the Verbal SAT®. ...

The test you are about to take is . . . especially difficult because [ST: we are interested in analyzing your abilities and limitations in reading and verbal domains so that we might better understand the factors involved] [RT: of our research focus on challenging reading tasks]. This test has previously been found valid for a wide range of populations including Caucasians and minorities, men and women, and individuals with and without learning disabilities, to name a few. Please provide strong effort in order to help us in our analysis of this problem-solving process. [RT: You will note that this test is different from almost all other tests that you have taken because we are not interested in analyzing ability: as was earlier mentioned, we are examining stylistic variations in problem-solving.] This test may also be helpful to you by familiarizing you with [ST: some of your strengths and weaknesses in reading and in verbal domains] [RT: the kinds of problems that appear on tests you may encounter in the future]. …

You will have 30 minutes to work on this test, which consists of 45 questions. … Questions left blank (because you felt unsure of the answer or because you ran out of time) will receive no credit or penalty: your score is based on only those questions you answer. It will, however, generally benefit you to guess if you are able to narrow a question down to two possible responses.