Journal of Experimental Child Psychology 220 (2022) 105417

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Effects of test difficulty messaging on academic cheating among middle school children

Li Zhao [a,*], Junjie Peng [a,1], Liyuzhi D. Dong [b], Yaxin Li [a], Haiying Mao [a], Brian J. Compton [c], Jiacheng Ye [a], Guoqiang Li [d,*], Gail D. Heyman [c], Kang Lee [b]

[a] *Department of Psychology, Hangzhou Normal University, Hangzhou 311121, People's Republic of China*
[b] *Dr. Eric Jackman Institute of Child Study, University of Toronto, Toronto, Ontario M5R 2X2, Canada*
[c] *Department of Psychology, University of California San Diego, La Jolla, CA 92093, USA*
[d] *Department of Primary Education, Hangzhou Normal University, Hangzhou 311121, People's Republic of China*

A R T I C L E   I N F O

Article history:
Received 23 September 2021
Revised 11 February 2022
Available online xxxx

Keywords:
Academic cheating
Self-scoring
Difficulty messaging
Approach achievement motivation
Avoidance achievement motivation
Moral behavior

A B S T R A C T

Academic cheating is a serious worldwide problem that begins during childhood. However, to date there has been little research on academic cheating with children before high school age. The current study used a naturalistic experimental paradigm to evaluate the possibility that systematically manipulating messages about the difficulty of a test can affect whether middle school children (N = 201) would cheat by reporting a falsely inflated test score. We found that test difficulty messaging significantly affected children's cheating behavior. Specifically, telling children that a test was either easy or hard produced higher rates of cheating than telling them that the difficulty level was on par with their current skills. In addition, among the children who chose to cheat, telling them that the test was easy led to a greater degree of cheating. These findings are consistent with theories of academic cheating that point to the importance of approach and avoidance motives in achievement motivation. The findings also suggest that simple messaging can have a significant impact on children's moral behavior and that seemingly innocuous messages such as describing the difficulty of a test can influence children's decisions about whether and how much to cheat.

* Corresponding authors.
E-mail addresses: zhaoli@hznu.edu.cn (L. Zhao), zxzxlgq@126.com (G. Li).
[1] L. Zhao and J. Peng are co-first authors.

https://doi.org/10.1016/j.jecp.2022.105417

Introduction

Academic cheating is a serious worldwide problem (Alan, Ertac, & Gumren, 2020; Anderman & Midgley, 2004; Callahan, 2004; Hrabak et al., 2004; Zhao, Zheng, Compton, Qin, & Heyman, 2020; Zhao et al., 2021). Most of the existing research in the field has focused on high school and university students (e.g., Flynn, Reichard, & Slane, 1987; McCabe & Bowers,1994; Waltzer & Dahl, 2021; Wenzel & Reinhard, 2020; Whitley, 1998; see Anderman & Murdock, 2011, and Cizek, 1999, for reviews). This research has found that the majority of students report having engaged in some form of academic cheating. The small number of studies that have been conducted with middle school children (e.g., Feldman & Feldman, 1967; Josephson Institute of Ethics, 2006; Keehn, 1956) suggests that about half of children have engaged in cheating by this age, but little is known about when and why they cheat. There is theoretical work suggesting that the social context in which cheating decisions are made is likely to be influential (Anderman & Murdock, 2011; Hartshorne & May, 1928; Murdock, Beauchamp, & Hinton, 2008; Murdock, Hale, & Weber, 2001), but there has been little experimental work on identifying specific contextual factors that affect academic cheating. The current research used a naturalistic experimental method to investigate the effects of one such contextual factor—messages about test difficulty.

Previous studies on the role of social contextual factors in academic cheating have largely used surveys (Baird, 1980; David, 2015; Davis, Grover, Becker, & McGregor, 1992; Ives et al., 2017; Ives & Giukin, 2020; McCabe, 1992). Although surveys have led to valuable insights about academic cheating and its development, they cannot establish whether specific situational factors actually cause children to cheat. Furthermore, due to the sensitive nature of cheating, children might not be motivated to provide honest answers to survey questions due to concerns about the possibility of being sanctioned for admitting to cheating (Erickson & Smith, 1974; Sideridis, Tsaousis, & Al Harbi, 2016; Zhao, Heyman, Chen, & Lee, 2017).

These limitations can be addressed with experimental methods that have been used to investigate cheating in preschool children (Evans, O'Connor, & Lee, 2018; Heyman, Zhao, Compton, & Lee, 2019; Zhao et al., 2020; Zhao et al., 2019). These studies have been inspired by nudge theory from behavioral economics, which posits that people's behavior can be influenced in predictable ways by simple and potentially subtle changes to their environment (Thaler & Sunstein, 2008). Nudges can take a physical form (e.g., placing traffic cones on a road to redirect traffic) as well as a verbal form (e.g., placing "drive carefully" signs on the side of a highway to reduce traffic accidents).

Previous developmental research using experimental methods to investigate children's cheating behavior in the context of games has revealed that both physical and verbal nudges can affect cheating rates among 4- and 5-year-olds (Fu, Heyman, Qian, Guo, & Lee, 2016; Heyman, Fu, Lin, Qian, & Lee, 2019; Zhao et al., 2017). For example, telling preschool children that they have a positive moral reputation (Fu et al., 2016) causes them to behave more honestly, but telling them that they are smart causes them to be more likely to cheat (Zhao et al., 2017). However, to date it is not clear what types of verbal messaging affect cheating in academic contexts.

Many scholars have argued that test difficulty is a key situational factor in relation to children's decisions about cheating (Brown, 1984; Bushway & Nash, 1977; Kibler, 1993; Steininger, Johnson, & Kirts, 1964). The level of difficulty has been proposed to influence cheating in two ways. First, cheating is sometimes viewed as a viable path to academic success when tests are too difficult (the *high-difficulty, high-cheating* hypothesis). On the other hand, poor performance on an easy test can lead children to be concerned about appearing foolish, so they may cheat to avoid this outcome (the *low-difficulty, high-cheating* hypothesis).

These two possibilities are consistent with a theoretical model of academic cheating by Murdock and Anderman (2006), which posits that students' decisions about cheating are strongly influenced by their achievement motivation (for a meta-analytic review, see Krou, Fong, & Hoff, 2021). When taking a test, students may have an approach motivation, in which they focus on the positive outcomes they want to obtain such as high academic standing or respect from teachers and classmates. They may also have an avoidance motivation, in which they focus on negative outcomes they want to avoid such as punishment from adults or appearing incompetent to their peers. According to the Murdock and Anderman (2006) model, both approach motivation and avoidance motivation can lead to increased cheating. Thus, when students are informed that a test is hard, they may be motivated to cheat on it to appear smart (the high-difficulty, high-cheating hypothesis), and when they are informed that a test is easy, they may be also motivated to cheat to avoid appearing incompetent (the low-difficulty, high-cheating hypothesis) (Day, Radosevich, & Chasteen, 2003; Elliot & Covington, 2001; Good & Shaw, 2021).

There is evidence in support of each of these hypotheses. For example, one study using survey methods found that children in Grades 7–9 (average age = 12.78 ± 0.87 years) were more likely to cheat when tests were difficult (Cheung, Wu, & Huang, 2016), but another study found that 19-year-old students were more likely to cheat when tests were easy (Simmons, 1963). It is possible that this discrepancy is simply due to methodological differences or age differences in the samples. However, it is also possible that both easy and difficult tests lead to increased cheating, perhaps for different reasons, as suggested by the Murdock and Anderman (2006) model. The current study took an experimental approach to test causal predictions derived from this model regarding the relation between test difficulty and cheating behavior.

In the current study, we gave Chinese middle school children a math test. All children took tests that were of the same difficulty level. We randomly assigned them to one of three conditions that differed only in terms of messaging about the test's difficulty: *grade level, easy,* or *hard.* In the hard condition, we told children that the test was very hard and its difficulty was above their grade's level. It is likely that children will interpret this type of instruction as an opportunity to demonstrate their ability, given that hard tests are often used in competitions, and that children as young as 6 years understand that success on a hard task is particularly indicative of high ability (Nicholls & Miller, 1983). In contrast, in the easy condition, we told children that the test was very easy and its difficulty was below their grade level. In doing so, we signaled to children that they had the opportunity to avoid appearing incompetent. In the grade level condition, we told children that the test was of a standard difficulty that was on par with their grade level. This condition was used as a baseline comparison for the other two conditions.

To assess children's cheating behavior, we used a self-scoring paradigm (Antion & Michael, 1983; see Cizek, 1999; Erickson & Smith, 1974; Fakouri, 1972; Flynn et al., 1987; Sherrill, Salisbury, Horowitz, & Friedman, 1971; Whitley, Nelson, & Jones, 1999). This paradigm requires children to take a test and then score it themselves. However, unbeknownst to the children, the experimenter takes photos of their test sheets before returning them for self-scoring, thereby allowing the researchers to identify any discrepancies between the children's self-reported performance and their actual performance.

The self-scoring paradigm is a well-established naturalistic method to study cheating in school (Antion & Michael, 1983; see Cizek, 1999; Erickson & Smith, 1974; Fakouri, 1972; Flynn et al., 1987; Sherrill et al., 1971; Whitley et al., 1999). This is because children are frequently asked to mark their own tests or homework and to submit the self-scored result. Specifically, this method is commonly used in China by teachers in all grades (Cai, 2010; Ministry of Education, 2012). The pedagogical rationale behind self-scoring is that it allows students to obtain timely feedback and provides them with an opportunity to see what they got wrong and learn from their mistakes (Cherepinsky, 2011; Edwards, 2007; Sadler & Good, 2006). However, one drawback is that some children may take advantage of the situation and surreptitiously correct their wrong answers to obtain a better grade. The self-scoring paradigm thus capitalizes on a common school practice to observe children's actual cheating behaviors in a naturalistic way (Antion & Michael, 1983; see Cizek, 1999; Erickson & Smith, 1974; Fakouri, 1972; Flynn et al., 1987; Sherrill et al., 1971; Whitley et al., 1999).

Based on the Murdock and Anderman (2006) model, we reasoned that if the high-difficulty, high-cheating hypothesis is correct and children are mainly motivated by approach motivation, they should be more likely to cheat in the hard condition than in the grade level condition. We also reasoned that if the low-difficulty, high-cheating hypothesis is correct and children are mainly motivated by the avoidance motivation, they should be more likely to cheat in the easy condition than in the grade level condition. We were also interested in examining whether cheating rates would differ between the hard condition and the easy condition as a way to assess whether children's cheating behavior is more strongly motivated by a desire to appear smart or by a desire to avoid appearing incompetent.

We had similar hypotheses about the extent to which children would falsely inflate their scores if they decided to cheat. To measure the extent of cheating, we obtained the difference between the scores provided by the children and their actual scores as determined by the experimenter. We tested whether the children who cheated would cheat more in the hard condition than in the grade level condition (the high-difficulty, high-cheating hypothesis), and whether they would cheat more in the easy condition than in the grade level condition (the low-difficulty, high-cheating hypothesis). We also investigated whether children who decided to cheat would cheat more in the hard condition than in the easy condition.

Method

Participants

We initially recruited a group of 204 children (97 girls) from eighth-grade classes at a middle school in Eastern China that is primarily composed of children from middle-class families. The final sample consisted of 201 children (94 girls) following the exclusion of 3 children who completed the exam but were absent from school when the self-scoring phase of the study was conducted 2 weeks later. All children were Han Chinese and had just started eighth grade (mean age = 13.36 years, *SD* = 0.49). Random assignment was done at the level of the classes, with two classes being randomly assigned to each of the three conditions, such that there were 66 children in the grade level condition (33 girls), 66 children in the easy condition (29 girls), and 69 children in the hard condition (32 girls). The study was approved by the university research ethics committee and was conducted as part of the participating school's educational research program. We obtained consent from children's parents or legal guardians.

Materials

We worked with the eighth-grade mathematics department of the middle school to develop three tests of equivalent difficulty based on workbooks that had not been used at the school before. Based on the teachers' suggestions, we selected items teachers typically use to assess whether students understand difficult mathematical concepts they have been taught. These items are typically constructed such that students will need to think carefully in order to get the right answers. This approach is intended to help students identify weaknesses in their understanding.

Because we were unable to test all classes at the same time, our design required three different tests of equivalent difficulty to ensure that children who had already taken the test would not be able to share information about it with children in different classes who had not taken it yet. To verify whether the versions of the test were indeed equivalent, we gave them to an additional eighth-grade class of 32 children and found that the mean scores on the three versions were not significantly different from each other (Version 1: M = 37.85, SD = 26.37; Version 2: M = 41.91, SD = 16.62; Version 3: M = 38.75, SD = 29.01), $F(2, 30) = 0.21$, p =.809, η_p^2 =.01.

Each version of the test contained 20 multiple-choice questions that were worth 3 points each and 10 fill-in-the-blank questions that were worth 4 points each for a total of 100 possible points. One version of the test is shown in the Appendix. In an effort to make the study as naturalistic as possible, the quantity and format of each type of question was based on the standard testing practices in the participating classrooms.

Procedure

The experimenter arrived at the classroom at the beginning of class and was introduced as a teacher. The children were told that the school board was conducting a school performance review and that, as a part of the review, each eighth-grade class would be taking a math test. They were further informed that different classes would be given tests of varying levels of difficulty, with one test that was easy for their grade level, one test that was difficult for their grade level, and one test that was appropriate for their grade level. As the experimenter distributed the tests, he informed the children of the difficulty level that had been assigned to their class. He also explained that there were colored stars at the top of each test sheet that indicated its difficulty, with one star indicating an easy test, two stars indicating a test that is appropriate for their grade level, and three stars indicating a hard test. These colored stars served to reinforce the difficulty messaging. Unbeknownst to the children, regardless of the star rating on their test, all classes took tests that were of the same level of difficulty.

Children were told that their class needed to complete the test within 20 min. They were also told that their test scores would be sent to their parents and would become an important part of their final math grade. This manipulation was included to ensure that the children would take the test seriously.

Following these instructions, the children completed the test using a black ballpoint pen under the supervision of the experimenter. When the 20-min period was over, the experimenter collected the completed test sheets, and after leaving the room he took a photo of each one. Two weeks later, he returned to the classroom for the self-scoring phase of the study.

The experimenter explained that he had been unable to find time to grade the tests and said the following:

> I was just told by the school board that I need to submit the scores today, so I am going to give the test sheets back to you, and you will have to mark your own test. Here are the grading instructions. Put a check mark by each answer that is correct and a cross mark by each answer that is incorrect. Then write your final score in the box that appears at the top of your test sheet. What you write here will be submitted as the score for your test. Please mark your test sheet according to the instructions. Do not make any attempts to tamper with your test sheet or score.

At this point, the experimenter reminded children of the difficulty level of their exam by once again explaining the meaning of the stars on their answer sheet. After ensuring that all students had received their own test sheet and instructing them not to communicate with any classmates, the experimenter used a projector to display the answer key at the front of the room. The experimenter gave the children 5 min to mark their own answer sheet using a red ballpoint pen. During this time, the experimenter pretended to take a phone call outside the classroom, where he remained visible to the children. After the 5 min, the experimenter returned and collected the self-scored test sheets.

Data coding and analysis

There were two dependent variables. The first was the presence of cheating, which was defined as a child self-reporting a score that was at least 1 point higher than his or her actual score, based on a review of the photos the experimenter had taken of each child's answer sheet prior to the self-scoring phase. The second was the cheating extent, which was computed by subtracting each child's actual score from his or her self-reported score. In addition, we also recorded the children's actual scores on the test, based on the photos that were taken of the test sheets prior to the self-scoring phase. We used SPSS 25.0 (IBM Corp., Armonk, NY, USA) for all of the following data analyses.

Coding was performed by two research assistants who did not participate in the study design or data collection phases and who were naive to the study's purpose. They first independently scored the two dependent variables. For the presence of cheating, the intercoder reliability was 100%. For the cheating extent, the initial inter-coder reliability was 95%. The initial inter-coder reliability for cheating extent was only 95% because several children cheated by correcting some of their answers and then awarding themselves a higher score than could be accounted for by the corrected answers. These inconsistencies were resolved by using the final reported score as the basis of the cheating extent measure.

We also coded specific forms of cheating into four mutually exclusive categories. The first category involved altering responses to multiple-choice questions, the second involved altering responses to fill-in-the-blank questions, the third involved altering both types of responses, and the fourth involved reporting an inflated final score without altering any responses. Children's responses were coded into these categories with an inter-coder reliability of 100%.

Presence of cheating

We used a binary logistic regression model to fit the presence of cheating responses, which was a dichotomous variable (no cheating = 0, cheating = 1). We then examined the main effect of our predictors, including condition (grade level, easy, or hard) and children's actual exam scores as the predicted variable. We first used the whole model with all predictors and their interactive terms, and we then selected the most parsimonious model following the convention of logistic analysis (Menard, 2001). We used the grade level as the reference condition for a priori comparisons.

Cheating extent

We conducted an analysis of variance (ANOVA) on the cheating extent scores, with condition (grade level, easy, or hard) as the predictor and grade level condition as the reference condition for a priori comparisons.

Results

Preliminary analyses showed no significant main effects or interactions involving the children's gender or classroom (*ps* >.05). Thus, the data were combined for these factors in all subsequent analyses.

Actual test scores

The mean actual score across conditions, which was coded based on the photos of the test sheets, was 46.09 (SD = 15.23, n = 201) out of 100. This value was 46.14 (SD = 15.66, n = 66) in the grade level condition, 45.35 (SD = 15.78, n = 66) in the easy condition, and 46.77 (SD = 14.47, n = 69) in the hard condition. These average scores were relatively low, as is consistent with our goal of avoiding ceiling effects. A one-way ANOVA revealed that the condition effect was not significant, $F(2, 198) = 0.15$, p =.87, η_p^2 =.001, Cohen's d = 0.06, which suggests that the three versions of the math test were comparable in terms of their actual difficulty.

Presence of cheating

As shown in Fig. 1, the cheating rate was 40.9% in the grade level condition, 62.1% in the easy condition, and 58.0% in the hard condition.

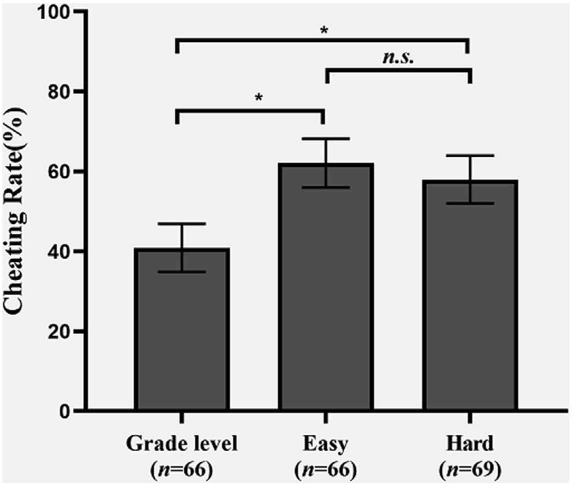


Fig. 1. Cheating rates by condition. Error bars indicate the standard errors, and asterisks indicate scores that significantly differed between groups in a priori contrasts or post hoc analyses. *n.s.*, nonsignificant; *p <.05.

Table 1
Percentage of children who engaged in each type of cheating by condition.

Condition	Multiple-choice questions only	Fill-in-the-blank questions only	Both types of questions	Final score only
Grade level (*n* = 27)	7.5% (2/27)	44.4% (12/27)	25.9% (7/27)	22.2% (6/27)
Easy (*n* = 41)	0% (0/41)	43.9% (18/41)	48.8% (20/41)	7.3% (3/41)
Hard (*n* = 40)	0% (0/40)	67.5% (27/40)	20.0% (8/40)	12.5% (5/40)
Total (*n* = 108)	1.9% (2/108)	52.8% (57/108)	32.4% (35/108)	12.9% (14/108)

Note. Numbers of children who cheated are shown in parentheses.

To evaluate the hypotheses relating to test difficulty, we conducted a binary logistic regression analysis using presence of cheating as the predicted variable. The most parsimonious model included only condition as the predictor. This model was significant, $\chi^2(2, N = 201) = 6.75$, p =.034, Nagelkerke R^2 =.044, −2LL (log likelihood) = 270.772. Inspection of the model revealed that the main effect of condition was significant, Wald $\chi^2(df = 2) = 6.62$, p =.037, Cohen's d = 0.37, which indicates that there were significant differences in the cheating rate under different test difficulty messaging conditions. A priori contrasts with grade level as the reference revealed that the cheating rate was significantly greater in the easy condition than in the grade level condition (62.1% vs. 40.9%; $\beta = 0.86$, SE = 0.36), Wald $\chi^2(df = 1) = 5.85$, $p < .05$, odds ratio (OR) = 2.37, 95% confidence interval (CI) = 1.18–4.76, Cohen's d = 0.48. In addition, the cheating rate in the hard condition was significantly greater than in the grade level condition (58.0% vs. 40.9%; $\beta = 0.69$, SE = 0.35), Wald $\chi^2(df = 1) = 3.89$, $p < .05$, OR = 1.99, 95% CI = 1.00–3.95, Cohen's d = 0.38. These findings support both the high-difficulty, high-cheating hypothesis and the low-difficulty, high-cheating hypothesis.

A follow-up post hoc comparison revealed that cheating rates in the easy and hard conditions were not significantly different (62.1% vs. 58.0%; $\beta = -0.17$, SE = 0.35), Wald $\chi^2(df = 1) = 0.24$, $p > .05$, OR = 0.84, 95% CI = 0.42–1.68, Cohen's d = 0.10, which suggests that children's tendency to cheat was motivated by either the approach or avoidance motivation.

Table 1 shows how children in the three conditions cheated. To test whether children who cheated varied in the methods they used to cheat, we conducted a chi-square analysis. We found a significant condition effect, $\chi^2(3, N = 108) = 65.11$, $p < .001$, Cohen's d = 2.46, which indicated that children were most likely to cheat by altering fill-in-the-blank questions only, or altering both multiple-choice and fill-in-the-blank questions.

We also performed four chi-square analyses to examine the condition effect for each of the four methods. We found that the condition effect was not significant for cheating by altering multiple-choice questions only, altering fill-in-the-blank questions only, or inflating the final score without altering any responses. However, children in the easy condition were more likely to cheat by altering both types of questions than children in the other two conditions, $\chi^2(2, N = 35) = 8.97$, p =.011. These results suggest that children who cheated used different methods to cheat, which was also affected by the condition to which they were assigned.

Cheating extent

Among the children who cheated, the cheating extent ranged from 1 to 61. As shown in Fig. 2, the cheating extent was greater in the easy condition (M = 16.29, SD = 13.45, n = 41) than in the grade level condition (M = 10.63, SD = 9.02, n = 27), and the cheating extent in the hard condition (M = 8.20, SD = 5.59, n = 40) was similar to that in the grade level condition.

To test the observed condition effect, we conducted an ANOVA on cheating extent, with condition as a between-subjects factor. The condition effect was significant, $F(2, 105) = 6.87$, p =.002, η_p^2 =.12, Cohen's d = 0.74. A priori contrasts with grade level as the reference revealed that among the children who cheated, those in the easy condition inflated their scores to a significantly greater extent than

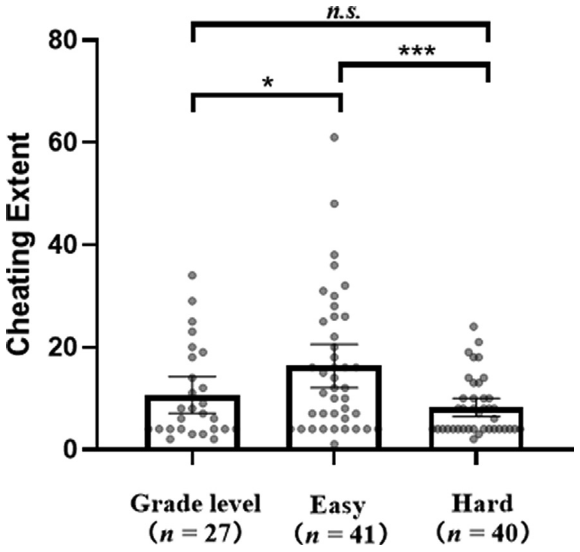


Fig. 2. Cheating extent among children who cheated by condition. Dots represent individual participants, and error bars indicate the standard errors, and asterisks indicate significant differences between conditions in post-hoc analyses. *n.s.*, nonsignificant; *** p<.001

their counterparts in the grade level condition (p =.025), but there was no significant difference between the hard condition and the grade level condition (p =.333). These results support the low-difficulty, high-cheating hypothesis but fail to support the high-difficulty, high-cheating hypothesis.

Post hoc analysis (least significant difference) revealed that among the children who cheated, those in the easy condition inflated their scores to a greater extent than children in the hard condition (p <.001), which suggests that once children decided to cheat, the extent to which they cheated was more strongly affected by avoidance motivation than by approach motivation.

Discussion

Children's academic cheating behavior has been theorized (Anderman & Murdock, 2011; Cizek, 1999) to be influenced by social contextual factors, one of which is verbal messaging. The current study used a naturalistic behavioral paradigm to examine whether middle school children's tendency to cheat on a test would be sensitive to messaging about its difficulty.

We obtained several major findings. First, averaged across conditions, more than half (53.7%) of the Chinese middle schoolers in our study cheated. This high rate of cheating is consistent with previous experimental and survey findings with middle school children in the United States. For example, using experimental methods, Feldman and Feldman (1967) found that 13-year-olds cheated at a rate of 28.4%, and Keehn (1956) obtained a cheating rate of 51.7% among children aged 11 and 13 years. Using survey methods, the Josephson Institute of Ethics (2006) found a cheating rate of 60% among American middle school students. These high rates of cheating among middle school students suggest that academic cheating is already well established at this age, perhaps due to competition among students and peer influences (Krou et al., 2021; Whitley, 1998). Consequently, researchers who are seeking to understand the emergence and early development of academic cheating will need to focus on children who are in middle school or are even younger. This approach should help us to identify empirically-based strategies to reduce cheating before it becomes normalized.

Second, we found that telling children that a test is either easy or hard made them more likely to cheat, as compared with telling them that the test's difficulty is appropriate for their grade level. This finding supports both the low-difficulty, high-cheating hypothesis and the high-difficulty, high-cheating hypothesis. It is also consistent with the predictions of the Murdock and Anderman (2006) model, which identifies three components of children's decisions about whether to cheat: goal, expectation, and cost. Regarding the goal, the model suggests that children can be motivated by an approach motivation (e.g., respect from teachers, parents, and peers) or an avoidance motivation (e.g., appearing incompetent, parental criticism, peer rejection). According to Cohen (1969), the size of the test messaging effect found in the current study is relatively small at 0.37, which means that other factors might have also contributed to children's decision to cheat. Nevertheless, our findings help to reconcile the apparently contradictory findings about test difficulty that have been found using survey methodologies (Cheung et al., 2016; Simmons, 1963) by suggesting that telling children that a test is easy can increase cheating by activating avoidance motivation, and telling children that a test is hard can increase cheating by activating approach motivation.

Studies on achievement motivation have revealed a number of motivational types that go beyond the simple dichotomy of approach versus avoidance orientations, including performance avoidance, performance approach, mastery avoidance, and mastery approach orientations (Daumiller & Janke, 2020; Elliot & McGregor, 2001). However, the way that these motivational patterns relate to academic cheating is unclear. For example, both Fakouri (1972) and Flynn et al. (1987) examined cheating among university students using a self-scoring method similar to that used in the current study. Fakouri (1972) found no link between cheating and motivation, and Flynn et al. (1987) found cheating to be linked to avoidance motivation but not approach motivation. Furthermore, a recent meta-analysis of studies involving high school and university students that primarily used survey-based methodologies (Krou et al., 2021) revealed that the mastery approach was negatively correlated with academic cheating and that cheating was not significantly associated with mastery avoidance, performance approach, or avoidance orientations. These inconsistencies could be due to factors such as whether academic cheating is being measured by experimental or survey methods, the instruments used to measure achievement motivations, or the age and educational level of the participants. Future research will need to systematically control for these factors.

Some factors besides motivational orientation that can influence children's cheating have been identified (for a meta-analytic review, see Krou et al., 2021). For example, children tend to cheat more if they think that a test is unfair (Cizek, 1999), if they want to avoid disappointing their parents (Jensen, Arnett, Feldman, & Cauffman, 2002), or if they think that their peers are cheating (for a meta-analytic review, see Anderman & Murdock, 2011). Cognitive skills (Evans & Lee, 2011), familial or school factors (Dahiya & Dahiya, 2019; McCabe, 2001), and dispositional factors such as personality (Burton,1963; Lobel & Levanon, 1988) can also make a difference. The roles of these factors could be explored in future research by using survey and interview methods.

Our third major finding is that among the children who altered their scores, all gave themselves additional points, and the tendency to inflate scores in this way was greater in the easy condition than in the grade level and hard conditions. The size of this effect was large, as indicated by a Cohen's d of 0.74, and it supports the low-difficulty, high-cheating hypothesis but not the high-difficulty, high-cheating hypothesis. This finding suggests that among the students who decided to cheat, a desire to avoid appearing incompetent (avoidance motivation) played a greater role than a desire to appear smart (approach motivation). However, it is also possible that the children who cheated inflated their scores strategically if they believed that adding points to an easy test would look less suspicious than adding points to a hard test. This suggestion is supported by the finding that children in the easy condition were significantly more likely to cheat by changing their answers on both types of questions as compared with children in the hard condition, who were more likely to cheat by changing their answers on the fill-in-the-blank questions only.

It should be noted that because the current study was conducted in China, several culture-specific or country-specific factors should be considered. For example, in China it is customary for teachers to keep parents apprised of their children's test scores, which might not be a standard practice in some other countries. In addition, Chinese culture is known for its emphasis on collectivism (e.g., Hofstede, 2001), which may play a role in mediating the relation between academic cheating and messages about test difficulty (e.g., students may be motivated to cheat by a desire to maintain grades that are similar to those of their peers). Future studies will need to explore these possibilities by testing children from multiple countries and cultures.

The current research has several limitations. The first concerns the difficulty of the test for students of this age. Although the test was designed to be difficult to avoid ceiling effects, such low scores could have unintended consequences. For example, children might have rejected the idea that the test was a good index of their ability, given that scores under 60% are considered a failing grade. Future studies should address this possibility by manipulating difficulty. A second limitation is that we did not obtain information about the reasoning behind children's decision making with reference to cheating, which could be addressed by performing confidential postexperiment interviews. For example, children could be asked to explain how they decided whether to cheat and whether the purported difficulty of the test influenced this decision. Third, the current study did not directly assess or manipulate children's academic motivations or their prior experiences with academic cheating. Future studies should conduct such assessments in addition to using behavioral paradigms to measure cheating (e.g., Antion & Michael, 1983; Erickson & Smith, 1974; Fakouri, 1972; Flynn et al., 1987; Midgley et al., 1998) to identify factors that can mediate the effects of test difficulty messaging on children's cheating behavior. Fourth, due to the naturalistic design of the current study, we did not tell children to refrain from discussing the test with students in other classes, and there might have been communication among students outside of the classroom setting that influenced our results. This could be addressed in future studies by testing each messaging condition in a different school.

The current study bridges a significant gap in the literature and advances our understanding of the development of academic cheating during childhood. It provides experimental evidence that academic cheating is already quite common by the middle school years. This finding, along with similar findings using survey methodologies, points to the need to study cheating before high school, when cheating behavior may have already become entrenched. In addition, the current study reveals, for the first time, that seemingly simple messaging about a test's difficulty can have a significant impact on children's cheating behavior during middle school, an age when children are highly sensitive to social information (Hanewald, 2013). The current findings show that information as seemingly innocuous as telling students how difficult a test is can influence their decision to cheat, and suggest that teachers may want to refrain from describing tests as either very easy or very hard.

In summary, the current study experimentally investigated the effects of messaging about test difficulty on middle school children's cheating, using a naturalistic behavioral paradigm. We found that test difficulty messaging significantly affected both the cheating rate and the degree of cheating. These findings, along with those of previous studies inspired by nudge theory (e.g., Zhao et al., 2017), suggest that simple messaging can have a significant impact on children's moral behavior and that seemingly innocuous messages such as describing a test as easy or hard may promote dishonesty.

Acknowledgments

The contributions of L. Zhao, J. Peng, Y. Li, and H. Mao were supported by the Zhejiang Provincial Philosophy and Social Science Planning "Young Talent Cultivation Project" of China (21QNYC18ZD) and the National Natural Science Foundation of China (31900773). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A

School Performance Tests (2020): Mathematics - Easy（I）

★☆☆

Total Score	

Notes for Students

1. This test examines the teaching quality at this school district. The test takes 20 minutes to complete.
2. All students must fill in their name and student number carefully according to the proctor's instructions. The student number is 2020 + class + student number, such as 20200101.
3. Write in black pen or pencil with legible handwriting.
4. Please complete the multiple-choice questions within the brackets, and complete the fill-in-the-blank questions within the given lines. Anything written outside of the designated answering areas will not be graded.

(Please fill in the following information carefully)

Student Number								
Student Name								

Part I. Multiple-Choice Questions (3 points for each question, 60 points in total)

1. Which of the following statements is true? ( ) [B]
   A. Ray PA and Ray AP are the same ray.
   B. Two points define a straight line.
   C. The length of Ray OA is 12 cm.
   D. The line AB and CD intersect at one point.

2. What is the reciprocal of negative rational number a? ( ) [B]

A. -a B. $\frac{1}{a}$ C. - $\frac{1}{a}$ D. a

3. Draw three straight lines on a plane. What is the maximum number of planes you can divide it into? ( ) [C]

A.4 B.6 C.7 D.8

4. Which of the following equations is correct? ( ) [B]

A. $(ab)^5 = ab^5$

B. $a^8 \div a^2 = a^6$

C. $(a^2)^3 = a^5$

D. $(a - b)^2 = a^2 - b^2$

5. Which of the following is a binary linear equation? ( ) [D]

A. x-y=3a B. $\frac{1}{3}$ - x=$\frac{1}{x}$ +2 C. -2a=3a+1 D. 2x-1=y

6. What is the answer to the following system of equations: x-y = 1, 2x + y = 5. ( ) [D]

A. x=-1, y=2 B. x=2, y=-1 C. x=1, y=2 D. x=2, y=1

7. Which of the following is the axis of symmetry of an isosceles triangle? ( ) [D]

A. the midline of the long base

B. the middle line of the leg

C. the bisector of the vertex angle

D. the vertical bisector of the base

8. Which of the following equations is remains correct after removing the brackets? ( ) [B]

A. a - (b - c) = a - b - c B. x^2-[-(-x+y)] = x^2- x+y

C. m-2(p - q) = m - 2p+q D. a+(b -c - 2d) = a+b - c+2d

9. First reduce the number *n* with 2, then multiply it by 5. W of the following is the final result? ( ) [B]

A. n-2×5 B.5(n-2) C. n-2+5n D.5n-2

10. Among -5, - $\frac{1}{10}$, -3.5, -0.01, -12, and -221, which has the greatest value? ( ) [C]

A. -12 B. - $\frac{1}{10}$ C. -0.01 D. -5

11. The distance between A and B is 100 km. Lily is at A, Carl is at B, and they are traveling toward each other. Lily's speed is 6km/h and Carl's speed is 4m/s. When the two meet, who is the closest to place A? ( ) [C]

A. Lily B. Carl C. they are the same distance from A D. unable to determine

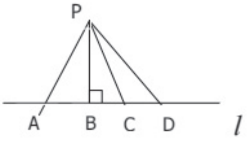
12. If m and n are natural numbers and m < n, what is the exponent of the polynomial $x^m + y^n - 2^{m+n}$? ( ) [B]

A. m-n B. n C. m D. m+n

13. As shown in the figure, which of the following indicates the distance from point P to line L? ( ) [A]

A. the length of segment PB B. the length of segment PA

C. the length of segment PC D. the length of segment PD



14. How many integers are greater than -7.1 and less than 1? ( ) [C]

A. 6 B. 7 C. 8 D.9

15. Assuming that 0＜a＜1, arrange a, $\frac{1}{a}$, and a^2 from least to greatest. ( ) [A]

A. a^2＜a＜$\frac{1}{a}$ B. a＜$\frac{1}{a}$＜a^2 C. $\frac{1}{a}$＜a＜a^2 D. a＜a^2＜$\frac{1}{a}$

16. When k = ______, the polynomial x^2-3kxy-xy-8 does not contain the term xy. ( ) [C]

A.0 B.$\frac{1}{3}$ C.$\frac{1}{9}$ D.- $\frac{1}{9}$

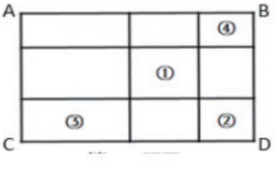
17. The exponent of 2^6-6x^3y^2+7x^2y^3-x^4-x is: ( ) [C]

A. 15 B. 6 C. 5 D. 4

18. When x = 1, qx^3+px+1=2017. When x = - 1, what is the value of qx^3+px+1? ( ) [B]

A. -2017 B. -2015 C. 2017 D. -2016

19. As shown in the figure, the rectangle ABCD is divided into nine quadrangles. Only ①and ② are squares, and the rests are rectangles. If the perimeter of square ①is 12, the perimeter of rectangle ③is 2m, the perimeter of rectangle ④is 2n, and 3 (m + n) + mn = 61, what is the area of this rectangle ABCD? ( ) [B]



A.60 B.70 C.80 D.90

20. If y=$|x+2|+|x-2|-|3x-6|$, what is the maximum value of y? ( ) [D]

A. 2 B. 3 C. 7 D. 5

Part II. Fill in the Blank（4 points for each question, 40 points in total）

1. The factors of x^2-4x^2 are ________. [$x^2(x-4)$]
2. In the equation x - 3y = 8, use x to represent y. y = ________. [$-\frac{8}{3}+\frac{x}{3}$]
3. In scientific notation -0.0000000459 = ________. [-4.59×10^{-8}]
4. If m is a positive real number and m - $\frac{1}{m}$ = 3, m^2 + $\frac{1}{m^2}$ = _______, and m^4 + $\frac{1}{m^4}$ = _______.

[11] [119]

5. The square root of 49 is ________. [±7]
6. The average speed of a ship when it travels from port A to port B is a, and when it travels from port B to port A its average speed is b. The average speed of this ship is _______. [$\frac{2ab}{a+b}$]
7. A cement factory’s annual output increases with a growth rate of 10%. If the output in the first year is a, the output in the third year is ________. [$\frac{121a}{100}$] .
8. If $|a| = 5$,$\sqrt{b^2} = 7$, and $|a+b|$ = a + b，then $|b-a|$ = ________. [2 or 12]

9. A man went shopping between 6 pm and 7 pm. He found that the angle between the hour hand and minute hand on his watch was 110° when he left, and it was also 110° when he returned. How many minutes was he away? _______. [40]
10. Given three primes a, b, and c, a + b+ abc = 99, so $|\mathbf{a}-\mathbf{b}|+|\mathbf{b}-\mathbf{a}|+|\mathbf{c}-\mathbf{a}|=$ _______. [34]

References

Alan, S., Ertac, S., & Gumren, M. (2020). Cheating and incentives in a performance context: Evidence from a field experiment on children. *Journal of Economic Behavior & Organization, 179*, 681–701.

Anderman, E. M., & Midgley, C. (2004). Changes in self-reported academic cheating across the transition from middle school to high school. *Contemporary Educational Psychology, 29*, 499–517.

Anderman, E. M., & Murdock, T. B. (2011). *Psychology of academic cheating*. Cambridge, MA: Academic Press.

Antion, D. L., & Michael, W. B. (1983). Short-term predictive validity of demographic, affective, personal, and cognitive variables in relation to two criterion measures of cheating behaviors. *Educational and Psychological Measurement, 43*, 467–482.

Baird, J. S. Jr., (1980). Current trends in college cheating. *Psychology in the Schools, 17*, 515–522.

Brown, D. (1984). *A comparison of student attitudes and perceptions regarding academic dishonesty of selected class groups in 1980 and 1983 at Iowa State University*. Doctoral dissertation, Iowa State University. doi:10.31274/rtd-180813-11833..

Burton, R. V. (1963). Generality of honesty reconsidered. *Psychological Review, 70*, 481–499.

Bushway, A., & Nash, W. R. (1977). School cheating behavior. *Review of Educational Research, 47*, 623–632.

Cai, J. (2010). *Basic teaching skills for middle school mathematics teachers [in Chinese]*. Beijing: Guangming Daily Press.

Callahan, D. (2004). *The cheating culture: Why more Americans are doing wrong to get ahead*. San Diego: Harcourt.

Cherepinsky, V. (2011). Self-reflective grading: Getting students to learn from their mistakes. *Primus, 21*, 294–301.

Cheung, H. Y., Wu, J., & Huang, Y. (2016). Why do Chinese students cheat? Initial findings based on the self-reports of high school students in China. *The Australian Educational Researcher, 43*, 245–271.

Cizek, G. J. (1999). *Cheating on tests: How to do it, detect it, and prevent it*. Mahwah, NJ: Lawrence Erlbaum.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. San Diego: Academic Press.

Dahiya, S., & Dahiya, R. (2019). Factors of academic cheating among secondary school students: Home environment and gender. *International Journal of Research in Social Sciences, 9*, 487–495.

Daumiller, M., & Janke, S. (2020). Effects of performance goals and social norms on academic dishonesty in a test. *British Journal of Educational Psychology, 90*, 537–559.

David, L. T. (2015). Academic cheating in college students: Relations among personal values, self-esteem and mastery. *Procedia—Social and Behavioral Sciences, 187*, 88–92.

Davis, S. F., Grover, C. A., Becker, A. H., & McGregor, L. N. (1992). Academic dishonesty: Prevalence, determinants, techniques, and punishments. *Teaching of Psychology, 19*, 16–20.

Day, E. A., Radosevich, D. J., & Chasteen, C. S. (2003). Construct- and criterion-related validity of four commonly used goal orientation instruments. *Contemporary Educational Psychology, 28*, 434–464.

Edwards, N. M. (2007). Student self-grading in social statistics. *College Teaching, 55*, 72–76.

Elliot, A. J., & Covington, M. V. (2001). Approach and avoidance motivation. *Educational Psychology Review, 13*, 73–92.

Elliot, A. J., & McGregor, H. A. (2001). A 2×2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501–519.

Erickson, M. L., & Smith, W. B. (1974). On the relationship between self-reported and actual deviance: An empirical test. *Humboldt Journal of Social Relations, 1*, 106–113.

Evans, A. D., & Lee, K. (2011). Verbal deception from late childhood to middle adolescence and its relation to executive functioning skills. *Developmental Psychology, 47*, 1108–1116.

Evans, A. D., O'Connor, A. M., & Lee, K. (2018). Verbalizing a commitment reduces cheating in young children. *Social Development, 27*, 87–94.

Fakouri, M. E. (1972). Achievement motivation and cheating. *Psychological Reports, 31*, 629–630.

Feldman, S. E., & Feldman, M. T. (1967). Transition of sex differences in cheating. *Psychological Reports, 20*, 957–958.

Flynn, S., Reichard, M., & Slane, S. (1987). Cheating as a function of task outcome and Machiavellianism. *Journal of Psychology, 121*, 423–427.

Fu, G., Heyman, G. D., Qian, M., Guo, T., & Lee, K. (2016). Young children with a positive reputation to maintain are less likely to cheat. *Developmental Science, 19*, 275–283.

Good, K., & Shaw, A. (2021). Achieving a good impression: Reputation management and performance goals. *Wiley Interdisciplinary Reviews: Cognitive Science, 12* e1552.

Hanewald, R. (2013). Transition between primary and secondary school: Why it is important and how it can be supported. *Australian Journal of Teacher Education, 38*, 62–74.

Hartshorne, H., & May, M. A. (1928). *Studies in the nature of character: Studies in deceit*. New York: Macmillan.

Heyman, G. D., Fu, G., Lin, J., Qian, M. K., & Lee, K. (2015). Eliciting promises from children reduces cheating. *Journal of Experimental Child Psychology, 139*, 242–248.

Heyman, G. D., Zhao, L., Compton, B. J., & Lee, K. (2019). Dishonesty in young children. In A. Bucciol & N. Montinari (Eds.), *Dishonesty in behavioral economics* (pp. 17–29). London: Elsevier Science & Technology Books.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage.
Hrabak, M., Vujaklija, A., Vodopivec, I., Hren, D., Marušić, M., & Marušić, A. (2004). Academic misconduct among medical students in a post-communist country. *Medical Education, 38*, 276–285.
Ives, B., Alama, M., Mosora, L. C., Mosora, M., Grosu-Radulescu, L., Clinciu, A. I., ... Dutu, A. (2017). Patterns and predictors of academic dishonesty in Romanian university students. *Higher Education, 74*, 815–831.
Ives, B., & Giukin, L. (2020). Patterns and predictors of academic dishonesty in Moldovan university students. *Journal of Academic Ethics, 18*, 71–88.
Jensen, L. A., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2002). It's wrong, but everybody does it: Academic dishonesty among high school and college students. *Contemporary Educational Psychology, 27*, 209–228.
Josephson Institute of Ethics (2006). *Josephson Institute report card on the ethics of American youth, 2006*. Los Angeles: Author.
Keehn, J. D. (1956). Unrealistic reporting as a function of extraverted neurosis. *Journal of Clinical Psychology, 12*, 61–63.
Kibler, W. L. (1993). Academic dishonesty. *NASPA Journal, 30*, 252–267.
Krou, M. R., Fong, C. J., & Hoff, M. A. (2021). Achievement motivation and academic dishonesty: A meta-analytic investigation. *Educational Psychology Review, 33*, 427–458.
Lobel, T. E., & Levanon, I. (1988). Self-esteem, need for approval, and cheating behavior in children. *Journal of Educational Psychology, 80*, 122–123.
McCabe, D. L. (1992). The influence of situational ethics on cheating among college students. *Sociological Inquiry, 62*, 365–374.
McCabe, D. (2001). Cheating: Why students do it and how we can help them stop. *American Educator, 25*, 38–43.
McCabe, D. L., & Bowers, W. J. (1994). Academic dishonesty among males in college: A thirty year perspective. *Journal of College Student Development, 35*, 5–10.
Menard, S. (2001). *Applied logistic regression analysis* (2nd ed.). London: Sage.
Midgley, C., Kaplan, A., Middleton, M., Maehr, M. L., Urdan, T., Anderman, L. H., ... Roeser, R. (1998). The development and validation of scales assessing students' achievement goal orientations. *Contemporary Educational Psychology, 23*, 113–131.
Ministry of Education (2012). *Teaching standards for compulsory education [in Chinese]*. Beijing: Beijing Normal University Press.
Murdock, T. B., & Anderman, E. M. (2006). Motivational perspectives on student cheating: Toward an integrated model of academic dishonesty. *Educational Psychologist, 41*, 129–145.
Murdock, T. B., Beauchamp, A. S., & Hinton, A. M. (2008). Predictors of cheating and cheating attributions: Does classroom context influence cheating and blame for cheating?. *European Journal of Psychology of Education, 23*, 477–492.
Murdock, T. B., Hale, N. M., & Weber, M. J. (2001). Predictors of cheating among early adolescents: Academic and social motivations. *Contemporary Educational Psychology, 26*, 96–115.
Nicholls, J. G., & Miller, A. T. (1983). The differentiation of the concepts of difficulty and ability. *Child Development, 54*, 951–959.
Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*, 1–31.
Sherrill, D., Salisbury, J. L., Horowitz, B., & Friedman, S. T. (1971). Classroom cheating: Consistent attitude, perceptions, and behavior. *American Educational Research Journal, 8*, 503–510.
Sideridis, G. D., Tsaousis, I., & Al Harbi, K. (2016). Predicting academic dishonesty on national examinations: The roles of gender, previous performance, examination center change, city change, and region change. *Ethics & Behavior, 26*, 215–237.
Simmons, N. (1963). *Cheating: A comparison of its incidence in self-scoring and on a paper and pencil test at Fort Hays Kansas State College.* Master's thesis, Fort Hays Kansas State College. https://scholars.fhsu.edu/theses/822..
Steininger, M., Johnson, R. E., & Kirts, D. K. (1964). Cheating on college examinations as a function of situationally aroused anxiety and hostility. *Journal of Educational Psychology, 55*, 317–324.
Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT: Yale University Press.
Waltzer, T., & Dahl, A. (2021). Students' perceptions and evaluations of plagiarism: Effects of text and context. *Journal of Moral Education, 50*, 436–451.
Wenzel, K., & Reinhard, M. A. (2020). Tests and academic cheating: Do learning tasks influence cheating by way of negative evaluations?. *Social Psychology of Education, 23*, 721–753.
Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education, 39*, 235–274.
Whitley, B. E., Nelson, A. B., & Jones, C. J. (1999). Gender differences in cheating attitudes and classroom cheating behavior: A meta-analysis. *Sex Roles, 41*, 657–680.
Zhao, L., Heyman, G. D., Chen, L., & Lee, K. (2017). Praising young children for being smart promotes cheating. *Psychological Science, 28*, 1868–1870.
Zhao, L., Heyman, G. D., Chen, L., Sun, W., Zhang, R., & Lee, K. (2019). Cheating in the name of others: Offering prosocial justifications promotes unethical behavior in young children. *Journal of Experimental Child Psychology, 177*, 187–196.
Zhao, L., Zheng, Y., Compton, B. J., Qin, W., & Heyman, G. D. (2020). The moral barrier effect: Real and imagined barriers can reduce cheating. *Proceedings of the National Academy of Sciences of the United States of America, 117*, 19101–19107.
Zhao, L., Zheng, J., Mao, H., Yu, X., Ye, J., Chen, H., ... Lee, K. (2021). Effects of trust and threat messaging on academic cheating: A field study. *Psychological Science, 32*, 735–742.
Zhao, L., Zheng, Y., Mao, H., Zheng, J., Compton, B. J., Fu, G., ... Lee, K. (2021). Using environmental nudges to reduce academic cheating in young children. *Developmental Science, 24* e13108.