Online Supplemental Material

Overview

This online supplement has the following information:

- Additional analyses and information about each of the five studies from the manuscript.
- Detail on the pilot study for the Study 2 intervention, assessing immediate effects on the meaningfulness of schoolwork.
- Additional text from the self-transcendent purpose intervention
- The full personal meaningfulness of schoolwork measure.

Study 1

Validity evidence for the diligence task. A large-scale validation study with two diverse samples of twelfth grade students from the Northeast US (total N = 826; Galla et al., 2014) indicated that the task demonstrated convergent validity with composite scores of self-control, grit, and conscientiousness (Pearson rs .14 to .24 for Study 1; rs .16 to .20 for Study 2, ps < .01), discriminant validity with unrelated constructs of extraversion and neuroticism (rs = -.02 to .02 for Study 1; rs = -.07 to .05 for Study 2; ps > .05), and incremental predictive validity by predicting senior-year GPA (β = .22 in Study 1 and β = .09 in Study 2) and standardized math scores (β = .12 in Study 1 and β = .16 in Study 2) after controlling for demographics (socioeconomic status, gender, ethnicity), IQ (fluid and crystallized intelligence), and baseline math fluency.

Validity of brief cognitive measures. In a sub-sample of n = 271 students in the present study at four schools, we had the opportunity to administer the full Kaufman Brief Intelligence Test (Kaufman, 1990) and the symmetry span task (Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004). These were significantly correlated with the present brief Raven's measure, rs = .40 and .32, ps < .001, respectively (and these correlations are not substantially different from the same correlations with the full Raven's measure; in a different pilot conducted by members of our research team with 100 participants from the Amazon Turk website [Buhrmester, Kwang, & Gosling, 2011], these correlations were r = .46 and .26, ps < .01, respectively).

Accounting for zero inflation in the diligence task. The behavioral measure—number of boring math problems solved correctly—was zero-inflated because many students only consumed tempting media, and some experts (J. Cohen, P. Cohen, West, & Aiken, 2003) have noted that in some cases OLS estimates can be distorted when predicting such measures. Therefore we repeated our analyses conducting a negative binomial regression. The same results emerged: a purpose for learning was a significant (p < .05) predictor of solving more problems, controlling for other motives for learning.

Study 2

Effectiveness of random assignment. Random assignment was effective. Participants in the purpose condition did not differ significantly from the control condition in terms of any of the three pre-intervention grading periods, t(331)=.47, *n.s.*, t(331)=.85, *n.s.*, t(331)=.90, *n.s.*, respectively. The conditions also did not differ in terms of gender, $\chi^2(1) = 1.18$, *n.s.*, racial

minority status (Hispanic / Latino or African American vs. White, Asian or Other) $\chi^2(1) = .00$, *n.s.*, placement in advanced math courses, $\chi^2(1) = 1.99$, *n.s.*, or age, $t(338) = 1.19$, *n.s.*

Study 3

Additional detail about the deeper learning procedure. For ethical reasons we could not make students' access to the test review contingent on their completing the intervention materials. Therefore, the survey was programmed so that students could skip any page of the intervention without writing any response; fortunately, no students did this.

In Study 3, the introductory material was clear that test review strategy that would lead to less learning would be to guess on the question right away and, if wrong, continue guessing quickly until arriving upon the correct answer by chance. If students used this strategy, time on each page would be lower on average. By contrast, a strategy that is more effective for learning and retaining the information is to think carefully about the question before trying to answer it, and, if wrong, returning to the source material or one's notes before trying a different answer. Students might also select an answer but then change the answer once or several times before submitting the page. If students used this strategy, time on each page would be higher on average. Of course, not all students are aware that the former strategy leads to greater learning, which is why students were given this information.

Effectiveness of random assignment. Randomization was effective. Purpose intervention students were no different from control students in terms of their scores on the first pre-intervention exam, $t(67) = 1.42$, $p = .16$, the second pre-intervention exam, $t(67) = 1.49$, $p = .14$.

Supplementary analyses of focal dependent variable. There were a number of methodological decisions to make when coding and analyzing the primary dependent variable—deeper learning behavior on the tedious test review. Here we examine what impact these decisions had on our primary conclusions.

First, we were aware of the possibility that students could take longer on each question because they were distracted or not paying attention; after all, students completed the tasks in naturalistic settings, presumably wherever they typically study. If this were true, then these distracted students would also be expected to need more guesses before identifying the correct answer to a question. There is no reason to expect that this would vary across conditions but it might have added error variance. Next, it was possible that higher-performing students should be able to answer more questions on their first guess, without much thought, and so their time per page might be lower on average, but only because they moved quickly through easier questions before spending more time on harder questions. To address these possibilities, the survey software also tracked the number of times participants selected a different response option on each page. Thus analyses could control for the average number of guesses per page so as to reduce variance in average time per page due to distraction or overall background knowledge of psychology. The number of guesses on each question did not differ by condition ($p > .3$). When conducting analyses controlling for average number of clicks on each page there was still a significant treatment effect, $b = .66$, $t(69) = 2.33$, $p = .023$.

Next, we re-conducted analyses when (a) treating as missing data any time measurement for a given page that was greater than 5 minutes, in order to reduce the possibility that our results were due to students leaving the task (e.g., to eat lunch), and (b) analyzing total number of seconds spent on the test review, as opposed to average time on each screen / question. These analyses supported the same overall conclusions (the main effects and interactions reported in the text were significant at $p < .05$ or marginally significant at $p < .06$).

Effects on final exam scores. Finally, recall that Study 3 was conducted in the context of a review for an exam. Although our primary dependent variable was deeper-learning behavior on this review exercise, it was also possible to explore whether the purpose intervention affected exam performance 1-2 days after the review activity. In an OLS regression predicting exam scores there was a positive but non-significant treatment effect in the full sample, $b = 2.21$ points out of 100, $t(69) = 1.02$, $p = .31$. This is the equivalent of going from an 85 to an 87 on the final exam.

Finding a significant treatment effect on an exam for *all* students, however, is a tall order for a brief intervention. In contrast to Study 2, which allowed for the possibility of many student actions over several months to add up to an overall effect on grades, Study 3 involved only one assignment—a test review that covered many weeks' worth of content. Students who spent many weeks not paying in attention or not coming to class—for instance, failing students—might have been seeing the content for the first time when they began the test review. If these students skipped the class meetings they would presumably not have any notes to review when completing the review activity. Among those students, perhaps all would benefit from the test review because they would be getting answers to questions they had never seen (the survey was programmed to not let them move on until they got the right answer). Hence the marginal benefit of the purpose intervention (about 30 seconds more per page) might be overshadowed by the overall effect of being given the answer to questions students had neglected throughout the semester. Consistent with this hypothesis, all low-performing students (i.e., students with a D average or below) improved their grade on the final exam, presumably as a result of the test review (Previous exam average: 67%; Final exam average: 78%, $t(18) = 5.43$, $p < .001$, $d = 1.25$; values are raw and unadjusted), and this did not vary by condition, $p = .38$.

By contrast, among the majority of students who showed at least some level of engagement with the learning material prior to the purpose intervention (students with an A to C- average), and who presumably had lecture notes and other material to review, the purpose intervention led to improvements in test scores compared to controls. Within this sub-sample of 51 students, a regression predicting final exam scores and controlling for prior exam scores showed a significant effect of the purpose intervention, $b = 4.66$ points on a 100-point scale, $t(49) = 2.45$, $p = .018$, $d = .57$. More concretely, students in the control condition showed no significant improvement from the average of their previous exams to the final exam (Previous exam average: 85%; Final exam score: 85%, $t(12) = 0.05$, $p = .96$, $d = .02$). In effect, control students, who were less likely to use a "deeper learning" strategy during the test review, did not show any learning gains compared to historical performance. But they did not do worse, which is important for for ethical reasons. By contrast, students in the purpose intervention condition showed significant gains on the final exam (Previous exam average: 82%; Final exam average:

88%, $t(37) = 4.69$, $p < .001$, $d = .76$). Having adopted a "deeper learning" strategy on the review, treated students actually learned more, it seems.

There are two important features of these findings to consider. First, the moderation by baseline performance does not mirror that presented in Study 2, which found that lower-performers benefitted most from a self-transcendent purpose. We believe this difference is more an artifact of our study design than it is support for a theoretical conclusion. That is, as shown above, all failing students benefitted dramatically from the test review. Unsurprisingly, a last-second test comprehensive review in a class they had put no effort into boosted their grade by a full ten percentage points. Among students who were at least minimally engaged in the course, however, there was a significant effect of a purpose, improving grades by nearly six percentage points in the purpose intervention and no improvement in the control. Second, it is important to acknowledge that although these moderation findings are theoretically interpretable and consistent with the overall account presented in the manuscript, they were not predicted. Therefore they should be treated cautiously until future studies replicate them. For these reasons, the findings from this exploratory analysis are reported in the online supplement and not emphasized in the main text.

Study 4

Effectiveness of random assignment. Random assignment was effective. In Sample 1, the experimental conditions did not differ in terms of age, $F(1, 116) = 0.46$, *n.s.*, gender, $\chi^2(1) = 0.14$, *n.s*, or final grade in psychology, $F(1, 103) = 0.01$, *n.s.* (recall that the experiment was conducted at the end of the term, after all assignments except one quiz had been administered, so the manipulation could not have affected cumulative grades). In Sample 2, the experimental conditions did not differ in terms of age, $F(2, 279) = .78$, *n.s.*, or gender, $\chi^2(2) = 0.05$, *n.s.* We were unable to acquire grades in the psychology course in Sample 2, but we did assess self-reported satisfaction with grades in college and found this did not differ by condition, $F(2, 282) = 1.33$, *n.s.*

Notes on data exclusions and final sample construction. Following recommendations for increasing the replicability of psychological science (Simmons, Nelson, & Simonsohn, 2011), here we outline in detail the rationale for the construction of our analytic sample and all data exclusions. We designed the study to be completed during daytime hours (7 A.M. to 10 P.M.) and not in the final hours in which studies could be completed for credit. This was done in order to prevent data collection from students who may be unable to concentrate on the study due to it being late at night or due to feeling rushed to complete the study in the final minutes in order to get credit (and perhaps participating in other online studies concurrently). For Sample 2, the survey software restricted access during nighttime hours and in the final hours before the subject pool was closed. Thus no data exclusions based on completion time were required. For Sample 1, however, that restriction was mistakenly not enforced (because the survey was not closed during the night and not ended in time on the final day of the study) and so some students participated in the late night or early morning hours or in the final two hours before the subject pool was closed ($n = 35$). We excluded these participants from analyses (and these numbers were not different by condition), in order to have comparable results across samples and to allow for a more theoretically interpretable test of our hypotheses.

Next, an important part of the task involves the opportunity to consume tempting media; however some students reported that their computers could not play the media correctly (n = 18), meaning it could not have tempted them away from the boring math problems. Because these students' data have a substantively different interpretation, they were also excluded from analyses. The effects of the purpose intervention compared to control were statistically significant when no data exclusions were carried out (i.e., N = 455; for math problems in Block 2, p = .03, for tempting media in Block 2, p = .02). (Note also that Simmons et al., 2011, recommend attempting precise replications when a study requires decisions about exclusion of data, something that we did with the two samples in Study 4).

Supplementary analyses. Here we report a number of additional analyses of data from the "diligence task." These support the same overall conclusion emphasized in the manuscript.

First, we found that the primary effect of the purpose intervention—a condition difference on Block 2—was significant when treating as missing data participants who failed to engage at all in Block 2, $t(276) = 2.88$, p = .004, rather than coding them as having completed zero problems as in the analysis presented above.

Next, we re-conducted analyses a negative binomial model rather than OLS, so as to account for the large number of zeroes in the dependent variable. When doing this there was again a significant main effect of the purpose intervention compared to control in Block 2.

Third, the same overall conclusion that a self-transcendent purpose improved self-regulation were supported when we analyzed the number of tempting videos or video games played, instead of number of math problems solved. In Block 2 there was a highly statistically significant difference between the purpose and control conditions, b = -.40, $t(281) = -2.88$, p= .004, d = .34, and this effect was significant in Sample 1, $t(99) = 2.06$, p = .04, d = .41, and Sample 2, $t(178) = -2.27$, p = .02, d = .34, and was no different across samples, interaction $t(276) = 0.03$, p = .98.

Pilot Study for Study 2: Assessing Immediate Effects of the Purpose Manipulation on Meaningfulness of Schoolwork

In the text of Study 2, we reported a pilot test of the effect of the purpose manipulation on the measure of the meaningfulness of schoolwork, which is a modified version of the Behavioral Identification Form (Vallacher & Wegner, 1989). While we report the contrast between the purpose intervention and the control in the text, this study also included a third, self-oriented condition that parallels Study 4 in the text. Here we report the full analyses of these data, as well as an exploratory moderation analysis.

Participants. Participants were N = 451 high school students from 13 different high schools across the country whose teachers asked to participate in our research. Nineteen percent of the students were Asian, 7% were African American, 28% were Hispanic / Latino, 29% were White, and 17% self-identified as "other." The high schools were diverse: eight were public district schools, four were public charter schools, and one was a private school. Schools varied widely in

socioeconomic characteristics; in five schools, 0-10% of students received free or reduced lunch because of their low household income; in two schools the figure was 11-50%; and in the remaining six schools, more than half of students were from low-income households. High school students were used as participants in the present study in order to speak more directly to the experimental effects on STEM-course GPA demonstrated in Study 2.

Procedure. Experimental procedures closely mirrored those in Study 2. The study consisted of two survey sessions spaced approximately two weeks apart (M = 13 days), some time in January or February. During the school day, teachers directed students to a website (www.perts.net) that delivered the two sessions. In the first session, participants completed a baseline survey assessing overall interest in school. In the second session, participants completed the manipulation materials, followed by an extended version of the meaningfulness of schoolwork items described in Study 1.

As noted, there were three experimental conditions: the control condition used in Study 2 in the text, the purpose condition used in Study 2 in the text, and a version of the self-oriented condition used on Study 4 in the text.

Considerable precaution was taken to prevent demand effects or experimenter influence. Teachers presented the experience as a survey, not an intervention; no mention was made to students that the purpose writing materials were designed to affect their thinking or behavior; and no teacher at the school had access to the materials or was aware of students' random assignment to condition. Researchers never interacted with students.

Measure of meaningfulness of schoolwork. Meaningfulness of schoolwork was measured immediately after the manipulation (see Study 1 for more detail; also see Figure 1 in the manuscript). We administered a 9-item measure that included additional behaviors that were similar to those described in Study 1 (e.g., copying math formulas off of the chalkboard). All of the items are reproduced at the end of this supplement. We summed across the items, with higher values corresponding to higher levels of construal (α = .73).

Measure of interest in schoolwork. To use in exploratory moderator analyses, we assessed students' intrinsic interest in school on the baseline survey (see Hulleman & Harackiewicz, 2009). We asked students to indicate how true the following two statements were about them: "I think school is interesting." and "I think what we're learning in classes is interesting." Participants answered both items on 7-point scales (Response options: *Strongly disagree, Disagree, Slightly disagree, Neither agree nor disagree, Slightly agree, Agree, Strongly Agree*). We averaged across these items, with higher values corresponding to greater interest (α = 80).

Results. First, we report the findings that also appear in the text. In an OLS regression with two dummy variables (one indicating the purpose condition and the self-oriented condition), the purpose manipulation led to greater meaningfulness of schoolwork compared to the control, $t(446) = 2.67$, $p = .007$, $d = .25$ (Control raw $M = 4.78$, $SD = 2.53$ vs. Purpose raw $M = 5.39$, $SD = 2.41$).

As our purpose for learning manipulation contained a large amount of self-oriented information that students were encouraged to connect to their self-transcendent aims, a secondary question was whether simply cueing students into thinking about intrinsic, self-oriented reasons for learning was sufficient to increase the meaningfulness of boring academic tasks. The self-oriented manipulation had no significant impact on students' judgments of the meaningfulness of their schoolwork relative to control, $t(446) = 1.21$, $p = .22$, $d = .11$ (Self-oriented raw $M = 5.05$, $SD = 2.41$).

The self-oriented condition scored between the purpose and neutral control, as in Study 4. Comparisons between the self-oriented and purpose manipulations revealed that students' levels of the meaningfulness of schoolwork in the two conditions did not significantly differ from one another, $t(446) = 1.20$, $p = .23$. This finding also mirrors Yeager et al.'s (2012) research, which showed that those with only a self-oriented motive for learning did not differ significantly from either group in terms of the meaningfulness of schoolwork**.**

Moderation by baseline interest in school. We explored whether the purpose manipulation might have the greatest effects on the meaningfulness of common school activities among students who reported that school overall was uninteresting. Consistent with our theory, in an OLS regression there was a significant Purpose manipulation (vs. neutral control) × Overall interest in school interaction, $b = -.37$, $t(451) = -2.71$, $p = .007$, $\beta = -.38$, such that uninterested students showed the greatest increases in level of the meaningfulness of schoolwork relative to controls. Although we conducted this test using a continuous variable, it is possible to illustrate this by conducting simple effects analyses within substantively meaningful sub-groups—that is, "interested" students (who "agreed" that school was interesting, i.e., average score over 5), and "bored" students (who did not "agree" that school was interesting i.e., average score of 5 or below). Simple effects analyses revealed that among "bored" students there was a significant effect of the self-transcendent purpose manipulation compared to controls, $b = 0.93$, $t(237) = 2.87$, $p = .004$, $d = .37$, but no effect compared to controls among "interested" students, $b = 0.06$, $t(214) = .17$, $p = .87$, $d = .02$. Thus, a purpose truly created meaning in the midst of boredom.

By contrast, the Self-oriented manipulation (vs. neutral control) × Overall interest in school interaction was non-significant, $b = -.17$, $t(451) = -0.90$, $p = .37$, $\beta = -.16$, and the self-oriented manipulation showed no effect compared to controls in either sub-group of baseline interest, $ps > .19$. However, simple effects analyses revealed a significant difference between the purpose and the self-oriented condition among "bored" students, $b = 1.07$, $t(237) = 3.04$, $p = .003$, $d = .39$, and not among "interested" students, $b = .05$, $t(214) = 0.13$, $p = .90$, $d = .02$. Thus, within the sub-group of "bored" students, the purpose condition differed significantly from the self-oriented group.

Summary. Altogether, the findings from this pilot study support the theoretical interpretation of the results reported in the manuscript. Specifically, (a) the purpose intervention altered the meaningfulness of schoolwork in the short term, as expected based on theory; (b) this was especially true among "bored" students, as expected based on theory; and (c) a highly similar self-oriented intervention was not sufficient to produce these benefits among disinterested students.

Additional Text from the Purpose Intervention

Below we reproduce the exact text from most of the self-transcendent purpose intervention.

In this part of the study, we are interested in learning more about your opinions about the world.
[page break]

How could the world be a better place? Sometimes the world isn't fair, and so everyone thinks it could be better in one way or another. Some people want there to be less hunger, some want less prejudice, and others want less violence or disease. Other people want lots of other changes. What are some ways that you think the world could be a better place? Everything you write will be kept anonymous.
[page break]

Thank you for sharing your thoughts. Next we would like to learn more about what kind of person you want to become.
[page break]

What do you want to get out of high school? What kind of person do you want to be?
A scientific survey of high school students by Stanford University

When you were younger, other people made all your decisions for you. You weren't in charge of your future. But in high school, you start to take more responsibility for yourself and for your future – to decide what to do and why to do it. One choice students make is how hard to work and what to work hard on. Almost everyone tries hard in school at least sometimes, and we're trying to learn why students put in that extra effort. So we started surveying high school students.

The first thing most students talked about is wanting to make more money by getting a good job. But we also learned that money was hardly ever the only reason—or even the main reason—when they really thought about it.

Most students said they also work hard because of unique and personal reasons, even though they usually don't talk about them to friends. Some students hadn't even put these reasons into words until they took the survey. But when they thought about it, it was these more personal reasons that really helped them get through the hard days when homework piled up and life got crazy. And it was these personal reasons that helped them see that even day—to—day activities had a lot of meaning for their life overall.

This is what students said motivated them to work hard:

- 64% said they wanted to be an educated person who has something to intelligent to say about what's going on in the world.
- 59% said they want to learn so they can make a positive contribution to the world.
- 78% said they want to gain knowledge so that they can have a career that they personally enjoy.
- 72% said they want to have the freedom to pick the life that they want to live.

[page break]

It's more interesting to hear this in students' own words. We picked three typical responses from normal students on the next few pages. Please read them, then we will ask if you could share your own thoughts.
[page break]

[Students read quotations from former students who had completed the materials previously, edited by researchers; contact the authors for access to these]
[page break]

Your turn: Why is learning important to your goals?

Our surveys are still in progress, and we would like students like you to be a part of the conversation.

Take a moment to think about what kind of person you want to be in the future. Also think about what kind of positive impact you want to have on the people around you or society in general. We're not asking about things like money, high status, or power – even though those things can be important. Instead, we want to know what knowledge and abilities you want to have, and why you want to have them.

Please take a few minutes to think about this and write your thoughts so they can be shared with others. Everything you write will be kept anonymous.

In the space below, write a few sentences that answer this question: *How will learning in school help you be the kind of person you want to be or help you make the kind of impact you want on the people around you or society in general?* (Don't worry about spelling or grammar. Just focus on getting your ideas across.)

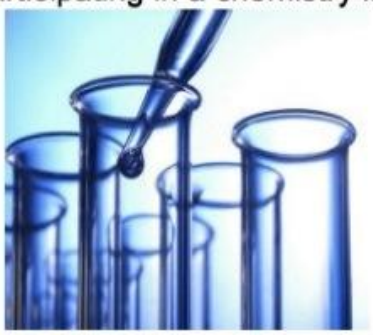
Personal Meaningfulness of Schoolwork Measure

Using a planner to record upcoming tasks

Choose the description that *more naturally comes to your mind* when you see the picture.

Staying organized so you can succeed in school and beyond. ○	Writing down different things you have to do in a little book. ○

Participating in a chemistry lab

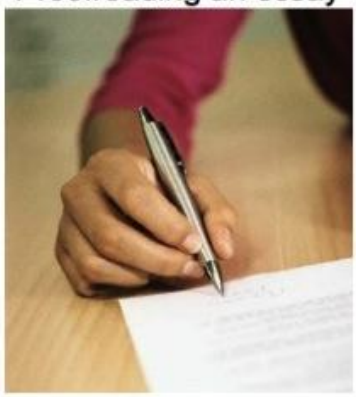


Choose the description that *more naturally comes to your mind* when you see the picture.

Learning how to conduct scientific studies.

Measuring different liquids in chemistry class.

Proofreading an essay

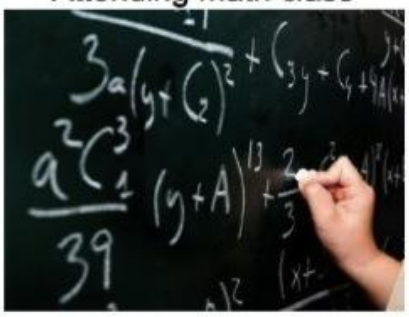


Choose the description that *more naturally comes to your mind* when you see the picture.

Making the changes to your essay that your teacher pointed out.

Learning how to express yourself more clearly.

Attending math class

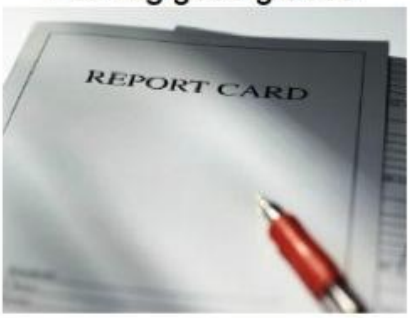


Choose the description that *more naturally comes to your mind* when you see the picture.

Copying notes off the board in math class.

Improving your ability to think critically about numbers.

Earning good grades

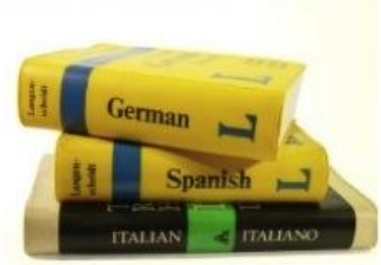


Choose the description that *more naturally comes to your mind* when you see the picture.

Getting a report card with As and Bs.

Gaining knowledge that will help you reach your goals in life.

Completing a homework assignment in a foreign language class.

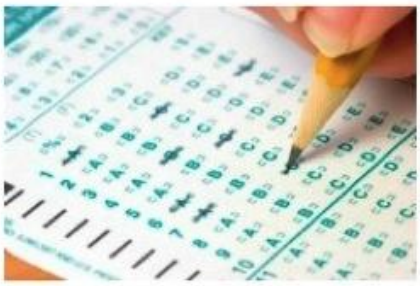


Choose the description that *more naturally comes to your mind* when you see the picture.

Preparing to communicate with people who speak a different language.

Reading words you don't understand and looking them up in a foreign language dictionary.

Taking the SAT



Choose the description that *more naturally comes to your mind* when you see the picture.

Taking steps towards a college degree.

Filling out the bubbles on the SAT.