How else could you do that? The effects of generating multiple means of goal attainment on female students' perceived goal attainability.

ABDULLA, A. and WOODS, R.

2022

Contemporary Educational Psychology 70 (2022) 102086

Contents lists available at ScienceDirect

Contemporary Educational Psychology

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

"How Else Could You Do that?" The effects of generating multiple means of goal attainment on female students' perceived goal attainability

Adam Abdulla *, Ruth Woods

Robert Gordon University

ARTICLE INFO

Keywords:
Perceived Goal Attainability
"Availability" Heuristic
Ease-of-Retrieval
Inconsistent Mediation

ABSTRACT

Perceived goal attainability (PGA) is a crucial variable in education, influencing students' goal commitment, goal pursuit and psychological wellbeing. Asking students to generate multiple means of goal attainment is thought to have a positive effect on PGA. And yet research on the "availability" heuristic suggests that difficulty in generating means of goal attainment may have a *negative* effect on PGA. The present study is the first to examine the matter in a real-world middle and high school context. In three experiments female students aged 11–15 were asked to generate many/few means of goal attainment. An inconsistent mediation model was hypothesised in which the "many means" condition has a negative indirect effect on perceived goal attainability through difficulty-in-generation (DIG) but a positive direct effect on the same variable. It was also hypothesised that these effects are greater in students with low baseline PGA. This moderated mediation hypothesis was supported statistically by tests of interaction in Experiments 1 and 2. In Experiment 3, which involved the youngest students with the highest baseline PGA, difficulty-in-generation and the "Think of many" manipulation appeared to have much less effect, again suggesting that the effect of DIG (and "Think of many") exert less of an influence when students' baseline PGA is high. Results have important implications for schools, students and educators alike.

1. Asking students to generate multiple means of goal attainment

Students are frequently encouraged to generate (multiple) means of goal attainment. For example, students are often asked in mathematics lessons "how *else*" they could solve a problem (Cianca, 2020; Parks, 2015; Sill & Smith, 2017). Students are also regularly asked how they could improve in various tasks such as reading/writing (Paratore & McCormack, 2005), note-taking (Jordan, 1997), or their overall learning (Evans & Brown, 2009). More generally, students are routinely invited to list "as many ways as possible" to attain a specific goal (e.g. Beghetto et al., 2015; Brusseau et al., 2020; Conklin, 2012; Gibbons, 2002; Lapp et al., 2011; Mazza et al., 2016; Stormont et al., 2012).

Educators ask students to generate multiple means of goal attainment for various reasons. One of the most prominent is the assumption that generating multiple means of goal attainment will enhance hope or (more specifically) perceived goal attainability, i.e. students' sense that a particular goal may be attained. It is assumed that the more means of goal attainment an individual identifies, the easier-to-attain the goal will appear. For example, Norrish (2015, p.242) asserts that "brainstorming multiple pathways to achieve the goal" is a key strategy for "building hope". Similarly, it is widely assumed that breaking down a challenging goal into simpler subgoals will "increase perceptions that the challenging goal [is] attainable" (Britt & Jex, 2015, p.90). Brophy (2010, pg.56) notes that proximal goals "usually seem attainable…even to students who doubt their capacities for attaining ultimate goals." Thus it would seem that asking a student to identify proximal goals or *means* of goal attainment should make the ultimate goal appear attainable as well. The importance of perceived attainability is now explained.

2. The importance of perceived goal attainability in education

Perceived goal attainability (PGA) is essential for students for multiple reasons. First, PGA is related to goal commitment. For example, a series of studies conducted with female secondary school students (involving a wide range of academic goals) found that lower/higher levels of PGA were associated with lower/higher levels of commitment (Abdulla & Woods, 2021a, 2021b, 2021c). Goal commitment is essential for goal attainment, particularly when goals are challenging (Klein et al., 2013). In other words, if students perceive a goal to be unattainable then

* Corresponding author.
E-mail address: a.abdulla@rgu.ac.uk (A. Abdulla).

https://doi.org/10.1016/j.cedpsych.2022.102086

Available online 28 June 2022

they are unlikely to make much effort to attain it. This in turn lowers their chances of success.

In three studies (with predominantly female participants) involving both mastery and performance goals, Senko and Hulleman (2013) found that the more difficult-to-attain a goal appeared (i.e. the lower students' PGA), the less likely students were to pursue it. In addition, low PGA was estimated to have negative indirect effects on students' level of academic interest and achievement.

PGA also has strong associations with wellbeing. For example, in another study with predominantly female participants, Gamble et al. (2020) found extremely strong links between (higher) PGA and (lower) depressive symptoms. Similarly, Brunstein (1993) found that when students had low PGA but remained committed to goals, their wellbeing was negatively affected. In other words, goal commitment – normally an advantage – may be detrimental without (high) PGA. The consequences are clear. In order to flourish academically and psychologically, students need to believe that their goals are attainable. Special attention should also be paid to PGA in *females*. Research suggests that although girls attach more weight (than boys) to educational goals they are generally less confident that they can attain them (Massey et al., 2009). Moreover, studies have found that expectations and perceived self-efficacy are lower in female students across a number of subjects including Maths (Huang, 2013; Mozahem et al., 2020; Watt, 2004). In addition, female students report greater anxiety regarding exams (Sung et al., 2016). This too may be related to PGA. For example, Atherton (2015) found that perceived likelihood of success in examinations was lower in female students. For all of these reasons some commentators have called for more focus on PGA in female students (e.g. Ringeisen et al., 2016; Watt, 2006). The present study is a response to that call. As already explained, educators may assume that inviting students to generate multiple means of goal attainment will enhance students' PGA. However, research on the use of the "Availability" heuristic suggests that it may have negative (as well as positive) effects.

3. The "Availability" heuristic and Difficulty-in-Generation (DIG)

Imagine a student – Gemma – whose goal is to improve her vocabulary. Her teacher therefore asks her to list a number of ways to achieve that goal. Suppose, however, that Gemma struggles to generate more than 1–2 strategies. That difficulty-in-generation (DIG) may actually *lower* perceived goal attainability (in Gemma). That is, the more difficulty students have in generating means of goal attainment, the more difficult-to-attain goals may appear. Support for this hypothesis derives from research on the "Availability" (or ease-of-retrieval) heuristic (Tversky & Kahneman, 1973). When relying on this heuristic, individuals form judgements on the basis of the ease/difficulty with which ideas come to mind. In a seminal study, participants asked to think of twelve examples of assertive behaviour - a difficult task - rated themselves as less assertive than participants asked to think of just six examples (Schwarz et al., 1991). Correlation analyses revealed that the more difficult it was to think of assertive behaviours, the lower people's self-reported assertiveness. This study provided researchers with an experimental paradigm hereafter referred to as "Think of many (vs. few)." It appears that in educational contexts students are also more heavily influenced by the ease/difficulty that they experience in generation than by the number of ideas that they generate. For example, Fuller et al. (2013) found that students asked to recall nine examples of academic success (a relatively difficult task) subsequently held *lower* opinions of their own academic ability than students asked to recall just three.

Although numerous "Think of many (vs. few)" studies have been conducted, questions have been raised about ease-of-retrieval effects (Replication Index, 2019). A *meta*-analysis conducted by Weingarten and Hutchinson (2018) found evidence of publication bias, potentially reducing the average effect size by a third. Moreover, researchers have failed to replicate the findings of Schwarz et al. (1991) when using a large random sample of adults (Yeager et al., 2019). Moreover, even if such effects are real in adults, much less is known about their operation in children. Geurten et al. (2015) found that young children (both male and female) asked to think of many first names (a difficult task) were subsequently more likely to conclude that they did not know many names than those asked to think of just a few. This study suggests that children may be just as likely as adults to rely on "availability" in the "Think of many" paradigm. Research also suggests that the perceived self-efficacy of middle and high school female students is influenced by ease-of-retrieval when students attempt to recall examples of success (Abdulla, 2021). However, only three published studies have investigated whether student PGA is influenced by "availability" in the context of generating means of goal attainment. Those studies are now briefly reviewed.

3.1. Previous research on generating means of goal Attainment, Difficulty-in-Generation (DIG) and perceived goal attainability (PGA)

Vaughn (1999) examined the effects of difficulty-in-generation (DIG) in the context of examinations. In one condition, college students were asked to list three things that might improve their chances of doing well on exams. In another, students were asked to list eight such things. Participants in the latter condition experienced considerably more DIG than those in the former. When reflecting on the "hardest" class exam at the beginning of the semester, those who had to think of eight things also judged themselves less likely to achieve A grades than those asked to think of just three. An effect size for the total effect of condition on PGA was not reported. However, a Cohen's d of 0.5 can be estimated from the data. This is an effect of medium size, according to Cohen's (1988) thresholds. Unfortunately, Vaughn (1999) provides no information about the gender of the students who participated.

Sanna and Schwarz (2004) conducted an experiment involving four experimental conditions - "3-success," "12-success," "3-failure," "12-failure" - and a control group. In the first two conditions, students were asked to list three/twelve things that might lead them to do well on an upcoming exam. One of the dependent variables was perceived likelihood of success (essentially, PGA), measured at two time-points: twenty-eight days in advance of the exam, and a few minutes before the exam. Students asked to list twelve things that might lead them to do well experienced greater DIG than students asked to list three such things. In addition, perceived likelihood of success was lower in the "12-success" condition (at both time-points). Indeed, the difference was greater than a point on the 0–10 scale, which suggests a large reduction in PGA induced by "Think of many." Like Vaughn (1999), Sanna and Schwarz (2004) provide no information about the gender of the students.

Finally, Chang (2010, Experiment 2) investigated whether difficulty in thinking of ways to treat a disease affects students' perceived ability to do so. Ninety-seven university students were randomly assigned to a "few ways' or "'many ways" condition. In the former, participants had to list three ways of preventing or treating haemorrhoids. In the latter condition, participants had to list seven ways of doing so. Participants in the "many ways" condition reported greater DIG and lower perceived ability to deal with the disease than those in the "few ways" condition. Chang (2010) did not report an effect size for the (total) effect of condition on PGA. However, a Cohen's d of approximately 0.42 can be calculated from the data provided. This effect is also (almost) of medium size. Chang (2010) notes that 47.8% of participants were male.

3.1.1. Previous research on DIG and PGA – Limitations and major problems

The aforementioned studies might appear to suggest that ease/difficulty in generating means of goal attainment has a positive/negative effect on students' perceived goal attainability. However, those studies have several limitations. First, two of the three studies provided no information about the gender of the participants. This makes it difficult (if

not impossible) to draw conclusions about female students. Second, the three reviewed studies involved undergraduates only. It is therefore not clear whether younger students are similarly influenced by "availability." On the one hand, it has been suggested that "heuristic responding should decrease with age, as the ability to override the heuristic response should improve over development" (Gualtieri & Denison, 2016, p. 2117). However, Gualtieri and Denison (2016, p.2117) describe this as a "largely untested assumption." On the other hand, some research suggests that as children grow older they are *more* likely to rely on heuristics (Furlan et al., 2013). For example, in an experiment involving children aged 6–11 and a comparison sample of undergraduates, Jacobs and Potenza (1991) found evidence to suggest that use of the "representativeness heuristic" increased with age, at least for social judgements. It therefore seems possible that the influence of "availability" on students' PGA will be stronger in older students.

However, there is a more serious limitation in the previous studies: the relationship between difficulty-in-generation (DIG) and perceived goal attainability (PGA) was never actually assessed. All three studies reviewed above assume that "Think of many" increases DIG, which then reduces PGA. That hypothesis is depicted in Fig. 1.

In the mediation model in Fig. 1, it is assumed that the *b* path is negative. That is, difficulty-in-generation (the putative mediator) is assumed to lower PGA. "Think of many" is then assumed to have a negative effect on PGA by increasing DIG. Vaughn (1999), Sanna and Schwarz (2004), and Chang (2010) all demonstrated that X ("Think of Many") causes M (Difficulty-in-Generation). In other words, they tested the *a* path. They also examined the total effect of X ("Think of Many") on Y (PGA). However, they did not assess the relationship between M and Y (i.e. DIG and PGA) - the *b* path. Without estimating that path it is impossible to test the mediational hypothesis. Moreover, it is impossible to determine the size of the effect of DIG on PGA, if indeed there is any.

4. "Think of Many Ways" and inconsistent mediation

Vaughn (1999), Sanna and Schwarz (2004), and Chang (2010) all reported a negative total effect of "Think of many" on PGA. However, this leaves the mechanisms of "Think of many" unexamined. In a mediation model such as the one in Fig. 1, the total effect of X on Y can be divided into an indirect effect through M, and a direct effect (independent of M). Importantly, the direct and indirect effects need not have the same sign. When one of these effects is positive and the other negative, methodologists speak of "inconsistent mediation" (e.g. MacKinnon et al., 2007). There are good reasons for supposing that the inconsistent mediation model applies to "Think of many."

Consider first the indirect effect in Fig. 1. This is the effect of "Think of Many" on PGA via difficulty-in-generation (DIG). This indirect effect can be calculated as the product of the *a* and *b* paths (i.e. *ab*). The *a* path is straightforward. Vaughn (1999), Sanna and Schwarz (2004) and Chang (2010) have all shown that "Think of many" raises DIG (relative to "Think of few"). If DIG subsequently has a negative effect on PGA, then the indirect effect of "Think of Many" on PGA (*ab*) will indeed be negative.

But now consider the *direct* effect (c'). This is the effect of "Think of Many" on PGA when difficulty-in-generation is held constant. All else being equal, it seems logical to suppose that the more means of goal attainment a student possesses, the easier-to-attain the goal will appear. Indeed, experimental research supports precisely that assumption (Huang & Zhang, 2013; Kruglanski et al., 2011). If so, then when DIG is held constant, "Think of many" should have a positive (direct) effect on students' PGA. Asking students to think of multiple ways to attain a goal may therefore have a negative indirect effect on PGA through DIG but a positive direct effect. If the negative indirect effect is large and the positive direct effect (fairly) small, then a total negative effect of medium-size may be found, as appears to have been the case in previous research (e.g. Vaughn, 1999). However, if the positive direct effect and negative indirect effect happen to be of similar size, they may cancel each other out, leading to a total effect close to zero. If so, then it would be misguided to look for effects of DIG by examining (only) the total effect. In the ease-of-retrieval literature, it has always been recognised that individuals may be influenced by both "availability" and "numerosity." However, it has generally been assumed that people are predominantly influenced by one or the other, depending on context (e.g. Schwarz et al., 1991; Tan & Agnew, 2016; Tormala et al., 2002). This "either..or" perspective obscures the fact that people may be *simultaneously* and perhaps equally influenced by both (albeit in opposite directions). Mediation analyses highlighting direct and indirect effects might shed important light on "Think of many" and ease-of-retrieval studies. Specifically, they might help to explain why famous findings have not been replicated (e.g. Yeager et al., 2019).

5. Moderation of the effects of DIG and "Think of Many" by baseline PGA

Another limitation in previous research on DIG and PGA is the failure to consider a major individual difference: students' *initial* PGA. And yet, there are good reasons for thinking that the effect of difficulty-in-generation on (posttest) PGA might depend on baseline PGA. Specifically, students in whom baseline PGA is relatively low may be more negatively affected by DIG than students in whom it is relatively high.

Weiner (1976) argued for a "low expectancy cycle": individuals low in expectancy (which is the same as PGA) are likely to attribute difficulties to stable factors, which lowers expectancy/PGA even further. On the other hand, those high in expectancy/PGA are likely to attribute difficulties to temporary factors, which means that their high PGA is maintained. Students in whom baseline PGA is low may therefore attribute DIG to a stable cause (e.g. a genuine lack of means), and develop even lower PGA as a result. On the other hand, students in whom baseline PGA is relatively high may attribute DIG to an aspect of the task or situation (e.g. the unusual request for so many means). If so, then their PGA may be largely unscathed. Empirical support for the sort of vicious cycle hypothesised by Weiner (1976) can be found in the literature on self-esteem. For example, research suggests that negative feedback has a more damaging impact on people low in self-esteem than on people high in self-esteem (Kernis et al., 1989). What applies to

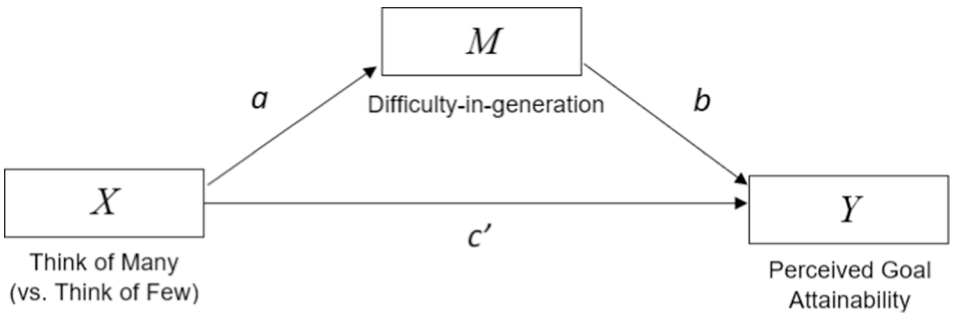

Fig. 1. Conceptual Mediation Model for the "Think of Many" vs. "Think of Few" Manipulation in the Context of Generating (Multiple) Means of Goal Attainment.

negative feedback and self-esteem may also apply to DIG and PGA.

More direct evidence is provided by Hermann et al. (2002), who actually examined the effects of "Think of many" on self-esteem. Students were asked to think of either two or eight examples of confidence-boosting experiences. Amongst those low in self-doubt (i.e. high in confidence), difficulty in thinking of examples (DIG) appeared to have little effect on self-esteem. However, among students high in self-doubt (i.e. low in confidence), DIG correlated negatively with posttest self-esteem. The researchers concluded that those high in self-doubt were "hypersensitive to retrieval difficulty compared to those low in self-doubt" (p.397). Students high in self-doubt are likely to be those with low baseline PGA. If (low) baseline PGA is a proxy for self-doubt, then one might expect a similar interaction, i.e. a negative association between difficulty-in-generation and posttest PGA when baseline PGA is low but not when it is high. In summary, high baseline PGA might act as a buffer against DIG whereas low baseline PGA might be a "risk factor." If so, then baseline PGA should moderate the effect of DIG on posttest PGA and (therefore) the indirect effect of condition (through DIG). Hermann et al. (2002) do not report the gender of their student participants. However, *female* students often report more self-doubt and lower (baseline) PGA than males (Atherton, 2015; Ringeisen et al., 2016). Female students may therefore be especially sensitive to DIG (as a result of greater self-doubt/lower baseline PGA).

In addition, the *direct* effect of "Think of many" may also depend on baseline PGA. It was suggested earlier that the direct effect of "Think of many" is likely to be positive. However, it may be more positive for those low in baseline PGA than for those who are already high. Students in whom baseline PGA is low may indeed benefit from generating means of goal attainment. For these students, the identification of multiple ways to attain a goal should make it seem easier to attain (Huang & Zhang, 2013; Kruglanski et al., 2011). On the other hand, students in whom PGA is already high may experience little additional benefit from generating multiple means. In their analysis of perceived self-efficacy (PSE), Gist and Mitchell (1992) suggest that there may be a certain fixity in those high in PSE such that there is unlikely to be much PSE improvement. On the other hand, low PSE is thought to be more malleable or plastic. Karl et al. (1993) report evidence supporting this "plasticity" hypothesis. In their study (in which 43% of participants were female) the provision of feedback had an impact on self-efficacy judgements for students low (but not for those high) in PSE. The PSE of students (initially) low in PSE "was more susceptible to influence from external cues than that of high self-efficacy individuals" (p.390). Perceived self-efficacy (PSE) is closely related to perceived goal attainability (PGA) and often measured by the same sorts of items (Klein et al., 2013). If low PSE is "plastic" (but high PSE is not) the same may be true of PGA. The generation of multiple means of goal attainment may then have a positive impact for students low (but not for those high) in baseline PGA. In short, baseline PGA might moderate the direct (as well as indirect) effect of "Think of many." The full moderated mediation model is depicted in Fig. 2.

6. The present study

In the present study, female students aged 11–15 were asked to generate many/few ways to attain a particular academic goal. They were asked to judge the likelihood of goal attainment both before and after the means-generation task. The age of the participants, number of ways and nature of the goal varied from experiment to experiment. The following hypotheses were of interest:

H1: "Think of many" increases DIG (relative to "Think of few").
H2(a): The effect of DIG on students' posttest PGA depends on baseline PGA.
H2(b): DIG has a more negative effect on students' posttest PGA when baseline PGA is low than when it is high.
H3(a): The direct effect of condition - "Think of many" (vs. "Think of few") – on students' posttest PGA depends on baseline PGA.
H3(b): The direct effect of condition on students' posttest PGA is more positive when baseline PGA is low than when it is high.
H4(a): The indirect effect of condition on students' posttest PGA depends on baseline PGA.
H4(b): The indirect effect of condition on students' posttest PGA is more negative when baseline PGA is low than when it is high.

Although developmental trends were not our primary focus, we considered the possibility that the effects of DIG might be larger in older students. Some previous research suggests that "between the ages of 7 and 15…children increasingly replace a mathematical decision-making model with a heuristic-based approach for judgement and decision making" (Lagattuta and Sayfan, 2013, p. 2095). Moreover, there are specific reasons for supposing that older students may be more likely to rely on "availability" (of means). Many studies have found that various forms of perceived self-efficacy (PSE) decline throughout middle/junior high school (Caprara et al., 2008; Schunk & Meece, 2006; Wigfield et al., 1991). For example, Fryer and Oga-Baldwin (2017) found that students' perceived self-efficacy declined in the first year of junior high school in both mathematics and foreign languages (two of the domains covered in the present study). Similar declines in PSE for mathematics have been observed by other researchers, especially in female students (e.g. Watt, 2004; Mozahem et al., 2020). Huang (2013) found no gender differences (in mathematics PSE) until the age of 15, by which point female students reported lower PSE for mathematics than males. Lower perceived self-efficacy means greater self-doubt. Greater self-doubt - often evident in females - apparently makes students more vulnerable to DIG (Hermann et al., 2002). Consequently, female students may be more influenced by difficulty-in-generation at 15 years old than they are at 11 or 12.

To enhance the generalisability of results, the focal goal varied from experiment to experiment as did the number of means-to-be-generated. Participants in the three experiments attended a private, all-female secondary school in London, UK. The school was attended by approximately 800 students (of differing ethnicities) ranging in age from 11 to 18, i.e. both middle and high school ages. Students were predominantly from high socio-economic backgrounds. The study was approved by the

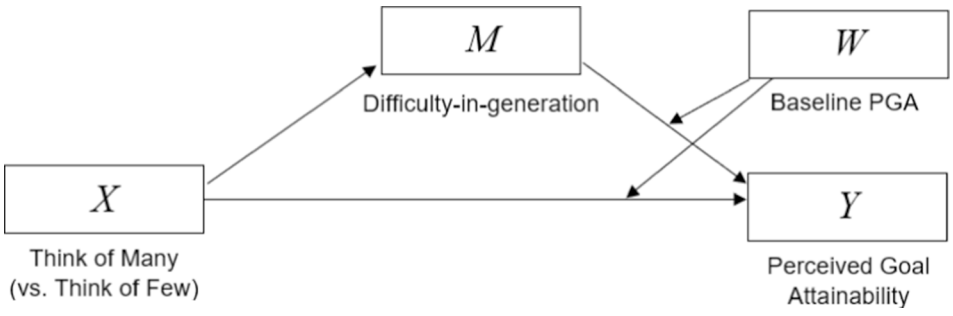

Fig. 2. A Moderated Mediation Model in Which the Indirect and Direct Effects of "Think of Many" on Posttest PGA are Moderated by Baseline PGA.

Ethics Committee at Robert Gordon University. The research was deemed to fall within the range of normal school activities, which meant that parental consent was not required (British Psychological Society, 2014). Participants themselves gave informed consent.

6.1. Sample size

In a mediation model such as the one in Fig. 1, power to detect an indirect effect (*ab*) depends on the size of both the a and b paths. It is extremely difficult to estimate the "required" number of participants when the size of these paths is unknown (Hayes, 2018). According to Fritz and MacKinnon (2007), if the *a* path is large, and the *b* path of medium-size, and if a percentile bootstrap test of mediation is applied, then 59 participants are required for 0.8 power to detect the indirect effect. Almost twice as many participants were recruited for Experiments 1 and 2 and more than twice as many for Experiment 3. This allowed not only for a b path of smaller than-predicted size but also for a test of moderation (Fig. 2).

6.2. Analytical strategy

A moderated mediation model was initially estimated in all three experiments. The independent variable was condition ("Think of many" vs "Think of few"); the mediator was DIG; the dependent variable was posttest PGA; and the moderator was baseline PGA. Baseline PGA was also included as a covariate in the equation for DIG. 95% bootstrapped confidence intervals were used for inferential purposes. Hayes' operationalisations for "low," "moderate" and "high" levels of a moderator were adopted. Thus, these were defined as the 16th, 50th and 84th percentile respectively.

Prior to analyses the data were inspected in order to identify extreme and potentially influential cases. This process involved examining histograms and boxplots and identifying the smallest and largest values of each variable as recommended by Darlington and Hayes (2017). After fitting the model, the data were again examined for extreme cases that might have undue influence on results. Judd et al. (2017) suggest that there is serious cause for concern when there are studentized deleted residuals with absolute values greater than 4 and noticeable gaps between the largest values of Cook's D. When a case was identified meeting at least one of these criteria, the analysis was conducted twice - once with, and once without that case. If results were essentially unchanged, the case was retained. If regression coefficients were noticeably affected by the extreme case, the analysis without that case was preferred (Judd et al., 2017).

In the present research, as in Vaughn's (1999) study, analyses were conducted twice - once including and once excluding students who did not list the full number of means. If the pattern of results was essentially the same, the analysis with all cases was considered to be supported. Finally, although *p* values and statistical significance are reported, more attention was paid to the direction and size of effects than to the extent to which a *p* value was less than 0.05.

6.3. Transparency and openness

We report how we determined our sample size, what manipulations were conducted, whether any data were excluded, and what measures were used. Quantitative data are available on request from the corresponding author. Data were analysed using SPSS and Hayes' PROCESS macro.

7. Experiment 1

Of the three published studies reviewed in the introduction, two used "Think of many" in the context of exams (Sanna & Schwarz, 2004; Vaughn, 1999). Putwain and Daly (2014) conducted a study in England involving 2435 students preparing for the General Certificate of Secondary Education (GCSE) in several academic subjects. They found that 16.4% of students reported experiencing exam anxiety "often" or "almost always." The proportion was higher in female students than in males (22.5% vs 10.3%). The focal goal in Experiment 1 was therefore "doing well in end-of-year exams" and participants were female students preparing for GCSEs.

7.1. Methods

7.1.1. Participants

One hundred and seventeen female students, all aged 14–15, were invited to take part in the experiment. Students were in their first year of their GCSE (General Certificate of Secondary Education) course. All of the invited students agreed to participate and completed both parts of the experiment. Internal (end-of-year) school exams were scheduled to take place three months after the experiment in several subjects. Participants were randomly assigned either to a "3 ways" (n = 59) or "7 ways" (n = 58) condition.

7.2. Procedure

The experiment was conducted remotely through Google Forms in two parts. Part 1 was the measurement of baseline PGA. Part 2 (a week later) was the manipulation. In Part 1, all participants were sent an email with a link to a Google Form. The Form stated the goal ("doing well in end-of-year exams") and then presented participants with the four items measuring baseline PGA. Participants completed the Form and then received a message indicating that Part 2 would take place in a week.

In Part 2, participants in each condition were sent an email with a link to the relevant Google Form. Students in both conditions were invited to consider how they could increase their chances of doing well in end-of-year exams. They were asked to write down one possible strategy in a box in the Form. The next two questions asked them "how else" they could achieve the goal (of doing well in the exams). On each occasion, they were asked to write down their idea in a box in the Form. At this point, participants in both conditions had been asked to generate three means of goal attainment. Participants in the "7 ways" condition were presented with the "how else?" question a further four times. After the means-generation questions, forms for both experimental conditions presented students with four items measuring posttest PGA, and four items regarding difficulty-in-generation.

7.3. Measures

7.3.1. Perceived goal attainability (PGA)

The perceived goal attainability (PGA) measure was based on a measure used in previous research with students in this age range (e.g. Abdulla & Woods, 2021a). The items on this measure are: "How likely is it that you will [focal goal entered here]?" "How easy will it be for you to [focal goal entered here]?" and "Imagine you had the goal of [focal goal entered here]. How 'achievable' would that goal be?" Student scores on this measure have repeatedly been found to correlate with scores on established measures of commitment (Abdulla & Woods, 2021a; 2021b; 2021c). Pilot studies indicated that reliability estimates were considerably enhanced when an additional item was included: "How confident are you that you will [focal goal entered here]?" Confirmatory factor analysis supported the assumption of unidimensionality for the resulting four-item measure. A 0–10 scale was presented for each item. Higher scores indicated higher PGA. Cronbach's alpha was extremely high at both baseline (α = 0.91) and posttest (α = 0.90).

7.3.2. Difficulty-in-generation (DIG)

This was assessed by a means of a four-item measure developed and tested in pilot studies. Participants were asked to provide a score between 1 and 7 for each of the following questions: "How hard was it to think of _ ways?" "How much effort did you have to make to think of _

ways?" "How challenging did you find it to think of _ ways?" "If you had to think of one more way, how difficult would that be?" The factorial validity of the measure was supported by confirmatory factor analysis, which indicated that the assumption of unidimensionality was reasonable. In pilot studies (as well as the reported experiments) DIG scores were consistently and reliably higher in "(Think of) Many" than in "(Think of) Few" conditions, which further supports the validity of the measure. Estimated reliability was excellent (α = 0.89). Higher scores indicated greater DIG.

7.4. Results

All participants in the "3 ways" condition were able to list three means of goal attainment. 93% of students (54 out of 58) in the "7 ways" condition were able to list 7 ways. The remainder (4 students) resorted to responses such as "I don't know" on at least one occasion. The analysis was conducted twice - once including and once excluding these participants. The pattern of results was the same in both cases. The inclusive analysis is reported below. Descriptive statistics for the key variables are displayed in Table 1. The highest baseline PGA score on the 0–10 scale was 7.5 and the highest posttest PGA score was 7.25.

One studentized deleted residual greater than 4 (4.12) was observed. This was also the case with the second highest value of Cook's D (0.42). The gap between this Cook's D and the third highest value was larger than the gap between the third and fourth, and fourth and fifth values. In order to clarify the influence of the case, the analysis was conducted twice - once with and once without the case – as recommended by Judd et al. (2017). The estimated effect of DIG was larger when the extreme case was included. In order to avoid bias, the extreme case was removed.

7.4.1. The effect of "Think of Many (vs. Few)" on DIG

With baseline PGA controlled, the effect of condition on DIG was large ($b = 0.82$ [0.43, 1.20], $t = 4.20$, $p = .0001$). Students in the "many" condition experienced far greater DIG than those in the "few" condition. H1 was clearly supported. The association between baseline PGA and DIG (controlling for condition) was very weak and not statistically significant (b = -0.07 [-0.21, 0.08], $t = 0.89$, $p = .37$). The standardized coefficient was −0.08 [-0.26, 0.10].

7.4.2. Moderation of the effect of DIG on Posttest PGA by baseline PGA

The interaction between DIG and baseline PGA in the model of posttest PGA was statistically significant ($b = 0.12$ [0.03, 0.20], $t = 2.80$, $p < .01$). The standardized coefficient was 0.15 [0.04, 0.26]. There was therefore support for H2(a), viz. that the effect of DIG on posttest PGA depends on baseline PGA. The Johnson-Neyman (JN) technique was used to probe the interaction. When baseline PGA was at or below approximately 5.32 (on the 0–10 scale), DIG was estimated to have a negative effect on posttest PGA. When baseline PGA was higher than 5.32 the effect of DIG was not statistically significant. Moreover, the lower the level of baseline PGA, the more negative the estimated effect. H2(b) was therefore supported. When baseline PGA was at 3.93 (the "low" level/16th percentile), the negative effect of DIG on posttest PGA was estimated to be −0.28 [- 0.44, −0.13].

Table 1
Means and Standard deviations for Measured Variables in Experiment 1.

	3 Ways		7 Ways	
	M	SD	M	SD
Baseline PGA	5.01	1.43	4.83	1.26
DIG	3.31	1.18	4.16	0.86
Posttest PGA	4.89	1.22	4.89	1.13

PGA = Perceived Goal Attainability.
DIG = Difficulty-in-Generation.

7.4.3. Moderation of the direct effect of condition on posttest PGA by baseline PGA

The interaction between condition and baseline PGA (in the model of posttest PGA) was right at the cut-off for statistical significance (b = -0.20 [−0.402, 0.004], $t = 1.95$, $p = .054$). The standardized coefficient was −0.23 [−0.460, 0.004]. There was therefore support for the hypothesis that the direct effect of condition on posttest PGA depends on baseline PGA - H3(a). Application of the JN technique indicated that when baseline PGA was below (approximately) 5.13, "Think of many" had a positive and statistically significant direct effect on posttest PGA. When baseline PGA was higher than 5.13, the direct effect of condition was not statistically significant. In addition, the lower the level of baseline PGA, the more positive the direct effect. Thus H3(b) was supported. When baseline PGA was at 3.93 (i.e. "low"), the direct effect of condition was estimated to be 0.51 [0.18, 0.84].

7.4.4. Moderation of the indirect effect of condition on posttest PGA by baseline PGA

The confidence interval for the index of moderated mediation did not include zero: 0.10 [0.003, 0.243]. There was therefore clear evidence to support the hypothesis that the indirect effect of condition depends on baseline PGA - H4(a). In addition, the lower the level of baseline PGA, the more negative the indirect effect. Thus, H4(b) was supported. When baseline PGA was at 3.93 (i.e. "low"), the negative indirect effect of "Think of many" (through DIG) was estimated to be −0.23 [−0.45, −0.04].

7.5. Brief discussion

Experiment 1 provided clear support for all hypotheses. The effect of difficulty-in-generation (DIG) on students' posttest perceived goal attainability (PGA) appeared to depend on baseline PGA. Posttest PGA was most influenced by "availability" of means (i.e. DIG) in students with lower baseline PGA. On the other hand, as predicted by the inconsistent mediation model, those with low baseline PGA were also the most likely to experience a *positive* direct effect of "Think of many." Vaughn (1999) and Sanna and Schwarz (2004) reported a negative total effect of "Think of many" on PGA in experiments involving undergraduates. However, they did not directly examine the association between DIG and PGA or assess direct and indirect effects. Nor did they test for moderation by baseline PGA. Moreover, they did not report the gender of their participants. Experiment 1 therefore adds considerably to their findings. First, direct evidence was provided for a negative effect of difficulty-in-generation on perceived goal attainability (at least among certain female students). Second, Experiment 1 indicated that susceptibility to DIG may be restricted to those with low baseline PGA. Those with high levels of PGA appear to be unaffected. Finally, Experiment 1 suggests that "accessibility" in "Think of many" influences PGA in female *high* school students. This in itself breaks new ground. Finally, Experiment 1 represents an advance in the study of DIG and PGA in another important respect. Previous studies (e.g. Sanna & Schwarz, 2004; Vaughn, 1999) have assumed that difficulty-ingeneration lowers PGA. But those studies did not include baseline PGA as a predictor. This leaves the association between "accessibility" (i.e. DIG) and PGA open to another interpretation. Specifically, it might be argued that difficulty-in-generation is associated with (posttest) PGA only because lower baseline PGA increases DIG. In other words, perhaps this is a case of reverse causation (with the effect running from PGA to DIG, rather than from DIG to PGA). By including baseline PGA in the model we were able to rule out that possbility: difficulty-in-generation predicts posttest PGA *even when baseline PGA is in the model.*

It is important to consider the size of the effects on the 0–10 PGA scale. At a low level of baseline PGA (3.93), the effect of DIG on posttest PGA was estimated to be almost one-third of a point and the negative indirect of "Think of many" was estimated to be almost a quarter of a point. These may not appear to be large effects. However, from a rational

perspective, *any* effect of "availability" on PGA may be considered important. Moreover, when baseline PGA was at a low level of 3.93, "Think of many" was estimated to have a positive direct effect of over half a point, which is approximately half a standard deviation. This is undoubtedly a meaningful effect.

Like previous studies of DIG and PGA, Experiment 1 focused on examinations. Researchers and practitioners might wonder whether PGA is susceptible to "availability" in other educational contexts. A demonstration that ease/difficulty-in-generation influences PGA for other types of goal would be an important extension of the present study. One of the main aims of Experiment 2 was to explore the generality of DIG's influence on PGA. The other major aim of Experiment 2 was to assess whether the results of Experiment 1 would be replicated in *younger* female students. Given that most research on heuristics has involved adults, evidence that PGA is affected by DIG in even younger children would be a significant contribution.

8. Experiment 2

The focal goal in Experiment 2 was mastering foreign language vocabulary. In England, Modern Foreign Languages (MFL) is a compulsory subject for children aged 11–14 (UK Government, n.d.). However, the perceived self-efficacy of students in MFL courses in UK secondary schools is generally low (Molway & Mutton, 2019). In addition, some research suggests that *female* students in the UK may be especially likely to underestimate their MFL abilities (e.g. Graham, 2004). This self-doubt might make female students' PGA for an MFL-related goal vulnerable to difficulty-in-generation. In addition, perceived competence and perceived likelihood of progress in modern languages appears to wane by the end of the first year of secondary school (e.g. Graham et al., 2016). This too might increase susceptibility to DIG. A considerable amount of research on learning foreign languages has focused on the acquisition of vocabulary (Andrä et al., 2020; de Groot and van Hell, 2005; Tseng et al., 2006; Wyra & Lawson, 2018). Experiment 2 was therefore conducted in the context of vocabulary acquisition. For the sake of generalisability, the number of means-to-be-generated was changed to two for the "few" and six for the "many" condition. Piloting indicated that this was sufficient to generate a large group-difference in DIG.

8.1. Methods

8.1.1. Participants

One hundred and seventeen female students, all aged 12–13, were initially recruited and were randomly assigned to a "2 ways" (n = 58) or "6 ways" (n = 59) condition. Two participants (one from each condition) opted out. Participants were in their second year of secondary school, studying various foreign languages including French, German, Spanish and Mandarin.

8.1.2. Procedure

The procedure was identical to that used in Experiment 1. However, the goal was now "mastering foreign language vocabulary." Once again, there were two parts to the experiment. In Part 1 baseline PGA was measured. In Part 2 - a week later – the manipulation was carried out.

8.2. Measures

8.2.1. Perceived goal attainability (PGA)

PGA was assessed with the measure used in Experiment 1. Cronbach's alpha was extremely high at both baseline (α = 0.90) and posttest (α = 0.92).

8.3. Difficulty-in-Generation (DIG)

DIG was assessed with the measure used in Experiment 1. Estimated reliability was again excellent (α = 0.92).

8.4. Results

All participants in the "2 ways" condition listed two means of goal attainment. 92% of participants (54 out of 59) in the "6 ways" condition were able to list six ways. The remaining participants resorted to responses such as "I don't know" on at least one occasion. The analysis was conducted twice - once with and once without these participants. The pattern of results was the same in each case. The inclusive analysis is reported below. Descriptive statistics for both conditions are presented in Table 2. The highest baseline PGA score was 9.75; the highest posttest PGA score was 9.25; and 97% of participants reported PGA scores lower than or equal to 8.5.

One exceptionally high studentized deleted residual was observed (-9.24), clearly exceeding the >4 cut-off recommended by Judd et al. (2017). This case also had the largest Cook's D (0.93). The gap between this value and the next highest value was far greater than the gap between any other values of Cook's D. The analysis was conducted once with and once without this case. Inclusion of the extreme case led to an unexpected (statistically significant) interaction between DIG and condition in the model of posttest PGA. There was no hint of an interaction when the extreme case was excluded, indicating that it had excessive influence. The case was therefore removed. All other cases were retained for the analysis.

8.4.1. The effect of "Think of Many (vs. Few)" on DIG

With baseline PGA controlled, the effect of condition was large: b = 1.14 [0.78, 1.51], t = 6.24 p <.0001. Students in the "many" condition reported considerably greater DIG than those in the "few" condition. H1 was therefore firmly supported. In addition, baseline PGA was negatively associated with DIG (controlling for condition): b = −0.31 [−0.46,−0.16], t = 4.16, p =.0001. The standardized coefficient was −0.14 [−0.27, −0.01].

8.4.2. Moderation of the effect of DIG on posttest PGA by baseline PGA

The interaction between DIG and baseline PGA in the model of posttest PGA was again statistically significant (b = 0.10 [0.02, 0.18], t = 2.45, p =.02). The standardized coefficent was 0.11 [0.02, 0.21]. There was therefore more evidence to suggest that the effect of DIG on posttest PGA depends on baseline PGA - H2(a). The JN technique was used to probe the interaction. Results suggested that when baseline PGA was at or below approximately 5.75, DIG had a statistically significant negative effect on PGA. Baseline PGA was lower than 5.75 in over 30% of participants. At levels of baseline PGA greater than 5.75, the "effect" of DIG was less negative and not statistically significant. Results therefore suggested that DIG has a more negative effect on posttest PGA when baseline PGA is low than when it is high - H2(b). In Experiment 2, the "low" value of baseline PGA was 5 (on the 0–10 scale). At this level of baseline PGA, the negative effect of DIG on posttest PGA was estimated to be −0.21 [-0.37, −0.05]. When baseline PGA was as low as the "low" level in Experiment 1 (i.e. 3.93), the estimated negative effect of DIG on PGA was −0.31 [−0.54, −0.09].

Table 2
Means and Standard Deviations for Measured Variables in Experiment 2.

	2 Ways		6 Ways	
	M	SD	M	SD
Baseline PGA	6.17	1.22	6.17	1.26
DIG	2.83	1.11	4.03	0.06
Posttest PGA	6.15	1.29	6.14	1.36

PGA = Perceived Goal Attainability.
DIG = Difficulty-in-Generation.

8.4.3. Moderation of the direct effect of condition on posttest PGA by baseline PGA

The coefficient for the interaction between condition and baseline PGA was right at the cut-off for statistical significance (b = -0.21 [−0.414, 0.002], t = 1.96, p =.053). The standardized coefficient was −0.19 [−0.388, 0.002]. There was therefore more evidence to suggest that the direct effect of condition on posttest PGA depends on baseline PGA - H3(a). Application of the JN technique indicated that when baseline PGA was below 5.68 there was a positive statistically significant direct effect of condition on posttest PGA. Over 30% of participants had baseline PGA scores under 5.68. When baseline PGA exceeded this value, however, the direct "effect" of condition was less positive and not statistically significant. Evidence therefore suggested that the direct effect of condition on posttest PGA is more positive when baseline PGA is low than when it is high - H3(b). When baseline PGA was at the "low" level of 5, the direct effect of "Think of many ways" was estimated to be +0.43, approximately half-a-point on the PGA scale. When baseline PGA was as low as the "low" level in Experiment 1 (i.e. 3.93), the positive direct effect was estimated to be as large as 0.66 [0.12, 1.20].

8.4.4. Moderation of the indirect effect of condition on posttest PGA by baseline PGA

The confidence interval for the index of moderated mediation was close to excluding zero: 0.11 [−0.01, 0.26]. There was therefore evidence (albeit not definitive) to suggest that the indirect effect of condition on posttest PGA depends on baseline PGA - H4(a). In addition, the indirect effect was estimated to be more negative at lower (than at higher) levels of baseline PGA, supporting H4(b). At the "low" level of 5, the negative indirect effect of "Think of many" (through DIG) was estimated to be −0.24 [−0.45, −0.07]. When baseline PGA was as low as the "low" level in Experiment 1 (i.e. 3.93), the negative indirect effect was estimated to be −0.36 [−0.72, −0.08].

8.5. Supplementary analysis

According to H2(a), the effect of DIG on posttest PGA depends on baseline PGA. The statistically significant interaction between DIG and baseline PGA was taken to support this hypothesis. However, in Experiment 2, baseline PGA and DIG were themselves correlated (r = −0.30, p <.01). When a focal variable (X) and putative moderator (W) are (highly) correlated, it can be difficult to distinguish between a linear interaction between X and W and a quadratic effect of either X or W (Aiken & West, 1991; Darlington & Hayes, 2017). If, for example, DIG has a curvilinear effect on posttest PGA (and if, as in the present case, DIG and baseline PGA are correlated), then an interaction between DIG and baseline PGA might be a "spurious moderator effect" (Lubinski & Humphreys, 1990).

One way to adjudicate between the moderator and quadratic effect hypotheses is to examine which model accounts for the most variance using stepwise regression techniques (Lubinski & Humphreys, 1990). MacCallum and Mar (1995) note that this effect size comparison approach can be reliable so long as the correlation between the predictors is less than or equal to 0.40, the reliabilities are greater than or equal to 0.70, and the sample size is not small. Those conditions were deemed to have been met in the present case. A stepwise regression analysis was therefore conducted. In the first step, posttest PGA was regressed on pretest PGA, condition, DIG, and the product of DIG and condition (i.e. all variables in our model except for the BaselinePGA × DIG interaction term). In the second step, DIG^2, BaselinePGA^2, and BaselinePGA × DIG were all entered in an incremental stepwise manner. Results of this analysis indicated that the addition of the BaselinePGA × DIG interaction term added most to the variance explained ($\Delta R^2 = 1.6\%$, p =.02). Neither quadratic term increased R^2 to a statistically significant degree. Results therefore supported the moderation (not the quadratic) hypothesis, according to which the effect of DIG on posttest PGA depends on baseline PGA - H2(a).

Fig. 3 depicts the differing effects of difficulty-in-generation (DIG) on posttest perceived goal attainability (PGA) at three different levels of baseline PGA. As may be observed, when baseline PGA is low, the downward slope indicates a negative effect of difficulty-in-generation. However when baseline PGA is moderate or high that negative effects appears to be absent.

8.6. Brief discussion

Experiment 2 yielded further support for all hypotheses. Difficulty-in-generation appeared to have a negative effect on perceived goal attainability at lower levels of baseline PGA. Similarly, "Think of many" appeared to have a negative indirect effect on posttest PGA when baseline PGA was relatively low but not when it was relatively high. The direct effect of condition was apparently also moderated by baseline PGA, although the sign of this effect (as predicted by the inconsistent mediation model) was positive rather than negative. In short, students appear to have been more sensitive to difficulty-in-generation and to the overall manipulation when initial PGA was low. Effect size estimates in Experiment 2 may be compared with those in Experiment 1. When baseline PGA in Experiment 2 was at the "low" level of Experiment 1 (i.e. 3.93 on the 0–10 scale), the negative effect of DIG on posttest PGA was estimated to be −0.31, which is very similar to the estimate in Experiment 1 (-0.28). At this "low" level of baseline PGA, the negative indirect effect of "Think of many" in Experiment 2 was estimated to be slightly larger than in Experiment 1 (−0.31 vs. −0.23), as was the positive direct effect (0.66 vs 0.51). However, these differences are small and it would seem more reasonable to speak of similarities.

However, Experiment 2 extended the findings of Experiment 1 in two important ways. First, Experiment 2 suggested that female students as young as 12 or 13 are influenced by "availability" (of means) when judging their likelihood of goal attainment. In other words, the results of Experiment 1 were replicated in a sample of even younger females. This is a significant finding, especially since the use of such heuristics by children has traditionally been questioned (e.g. Baron et al., 1993). Second, Experiment 2 indicated that difficulty-in-generation has a negative effect on the perceived goal attainability of a different type of goal - mastering foreign language. This too breaks new ground. Interestingly, effect size estimates suggest that DIG's influence in this context was as great as in the context of doing well in exams (Experiment 1).

Nevertheless, academic goals come in many other varieties. Readers may be wondered whether the findings of Experiment 2 extend to goals in other academic domains, or to goals in specific subjects such as mathetmatics. Research indicates that students' self-efficacy judgements differ across subjects and are associated with different developmental trajectories (e.g. Anderman & Midgley, 1997). Bong (2001, p.23) notes that despite correlations across subjects, students' "self-concepts are clearly divided along the line of verbal and math domains." It is therefore important to determine whether the findings of Experiment 2 (which involved a verbal goal) extend to goals relating to maths. One of the primary aims of Experiment 3 was to address this question. Another major aim was to see whether the findings of Experiments 1 and 2 would be replicated in a sample of even *younger* female students at a crucial point in their educational career- the transition from primary (elementary) to secondary (middle) school.

9. Experiment 3

Participants in Experiment 3 were in their very first year of secondary school, which is equivalent to the beginning of middle school in the US. They were the youngest students in the present study and the youngest students in their school. The focal goal was "improving in Maths in the remainder of the year". The transition from primary (elementary) to secondary (middle/junior high) school is widely regarded as critical. As children move from one school to another, they re-evaluate their academic abilities (Anderman & Midgley, 1997).

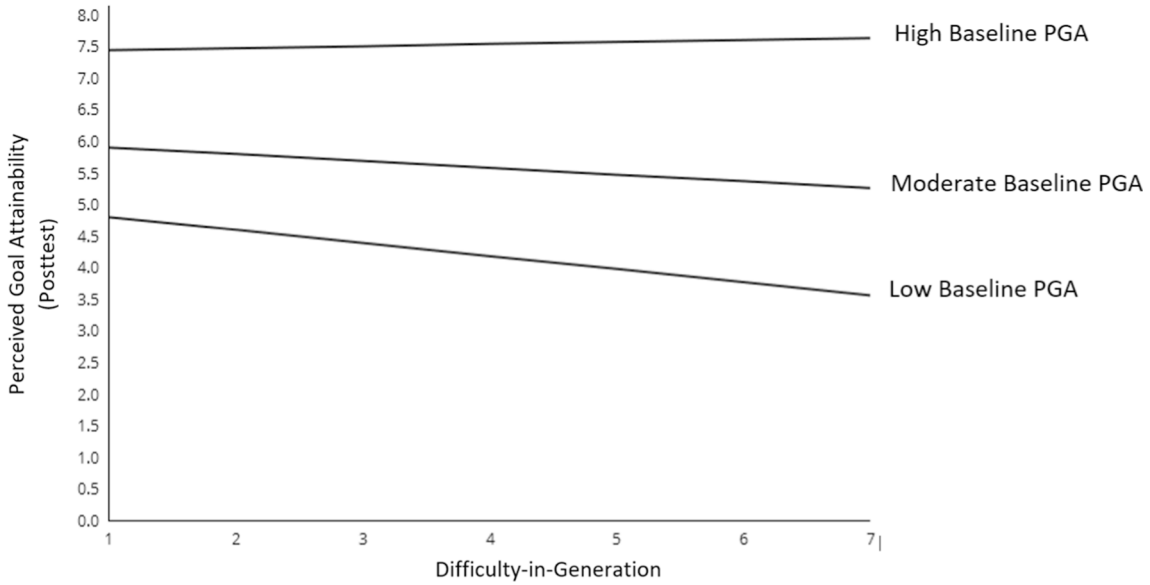

Fig. 3. The Effects of DIG on Posttest PGA at Low (16th percentile), Moderate (50th percentile) and High (84th Percentile) Levels of Baseline PGA.

Understanding whether PGA is influenced by "availability" at this stage is therefore extremely important.

Perceived competence in mathematics appears to decline in middle school, particularly amongst female students (Falco, 2019; Watt, 2004). However, in the very first year of secondary school (or middle school in the US), perceived self-efficacy (PSE) in mathematics appears to be relatively high and females feel no less confident in their mathematical abilities than males (Huang, 2013). This (relatively) high mathematics PSE should mean (relatively) high baseline PGA with regard to a maths-related goal. According to the present study's hypotheses, high(er) baseline PGA entails reduced susceptibility to difficulty-in-generation and to the "Think of many" manipulation. One might therefore predict that female students in the first year of secondary school are less affected by DIG and "Think of many" than older (female) students. Experiment 3 enabled us to test this prediction.

9.1. Methods

9.1.1. Participants

One hundred and forty-two female students, all aged 11–12, were invited to take part in the study. For all of these students Maths was a compulsory National Curriculum subject (UK Government, n.d.). Four students opted out. Participants were randomly assigned to a "3 ways" (n = 70) or "8 ways" (n = 68) condition. The number of ways (three and eight) was derived from Vaughn (1999).

9.1.2. Procedure

The procedure was identical to that used in Experiments 1 and 2. The experiment once again consisted of two parts: Part 1 (the measurement of baseline PGA), and Part 2 (the manipulation). Part 2 took place one week after Part 1, as before. The specified goal was "improving in Maths in the rest of the year.".

9.2. Measures

9.2.1. Perceived goal attainability

PGA was assessed by means of the measure used in Experiments 1 and 2. Cronbach's alpha was extremely high at both baseline ($\alpha = 0.91$) and posttest ($\alpha = 0.90$).

9.2.2. Difficulty-in-Generation

DIG was assessed by means of the measure used in Experiments 1 and 2. Estimated reliability was again excellent ($\alpha = 0.89$).

9.3. Results

Ninety-nine percent of participants (69 out of 70) in the "3 ways" condition were able to list three means of goal attainment. Seventy-eight percent of participants (53 out of 68) in the "8 ways" condition were able to list eight ways. The remaining participants resorted to responses such as "I don't know" on at least one occasion. The analysis was conducted twice - once including and once excluding these participants. The pattern of results was the same in each case. The inclusive analysis is reported below. Descriptive statistics for both conditions are presented in Table 3. One participant reported PGA of 10 (on the 0–10 scale) at both baseline and posttest. However, 91% of participants reported PGA scores lower than or equal to 8.5.

There were no studentized deleted residuals above 4 and the gap between the largest values of Cook's D was small (0.1). All cases were therefore included in the analysis.

9.3.1. The effect of "Think of Many" vs. "Think of Few" on DIG

With baseline PGA controlled, the effect of "Think of many" (vs. "Think of few") was large ($b = 0.94$ [0.62, 1.26], t = 5.75, p <.0001). Students in the "many" condition reported considerably more DIG than those in the "few" condition. H1 therefore received strong support. In addition, with condition controlled, baseline PGA was negatively associated with DIG ($b = -0.14$ [−0.27, −0.01], t = 2.16, $p = .03$). The standardized coefficient was −0.17[−0.32, −0.01].

9.3.2. Moderation of the effect of DIG on posttest PGA by baseline PGA

In the model of posttest PGA, the coefficient for the interaction between DIG and baseline PGA was small and not statistically significant

Table 3
Means and Standard Deviations for Measured Variables in Experiment 3.

	3 Ways		8 Ways	
	M	SD	M	SD
Baseline PGA	6.72	1.37	6.46	1.27
DIG	3.77	0.98	4.74	1.01
Posttest PGA	6.81	1.41	6.49	1.25

PGA = Perceived Goal Attainability.
DIG = Difficulty-in-Generation.

($b = -0.05$ [−0.14, 0.05], $t = 0.98$, $p = .33$). The standardized coefficient was also −0.05 [−0.15, 0.05]. There was therefore little statistical evidence to suggest that the effect of DIG on posttest PGA depends on baseline PGA. On the face of it, therefore, H2(a) and H2(b) were not supported. However, there is a clear sense in which they *were* supported, as explained in the discussion.

9.3.3. Moderation of the direct effect of condition on posttest PGA byBbaseline PGA

The coefficient for the interaction between condition and baseline PGA was also small and not statistically significant ($b = -0.07$ [−0.30, 0.16], t = 0.59, $p = .55$). The standardized coefficient was also −0.07 [−0.29, 0.16]. There was therefore apparently no strong statistical evidence to suggest that the direct effect of condition depends on baseline PGA. H3(a) and H3(b) were therefore apparently not supported - but, again, see the discussion.

9.3.4. Moderation of the indirect effect of condition on posttest PGA by baseline PGA

The index of moderated mediation was close to zero and the confidence interval also included zero: −0.04 [−0.15, 0.02]. There was therefore little statistical evidence to suggest that the indirect effect of condition on posttest PGA depends on baseline PGA. H4(a) and H4(b) were on the face of it not supported – but, once again, see the discussion.

Both interaction terms were dropped and a simple mediation model was estimated. The effect of DIG on posttest PGA (controlling for baseline PGA) was estimated to be negative and the coefficient was almost statistically significant ($b = -0.12$ [−0.26, 0.01], t = 1.79, $p = .08$). The standardized coefficient was −0.10 [−0.22, 0.01]. The indirect effect of condition on posttest PGA was also estimated to be negative (−0.12). The confidence interval was extremely close to excluding zero: [−0.272, 0.009]. The partially standardised estimated effect size for the indirect effect was −0.09 [−0.208, 0.006]. The estimated direct effect of condition on posttest PGA was positive but extremely small and not statistically significant ($b = 0.006$ [−0.29, 0.30], t = 0.04, $p = .97$). The total effect was estimated to be negative (−0.11) but was not statistically significant (t = 0.81, $p = .42$). Finally, there was an extremely strong association between baseline PGA and posttest PGA ($b = 0.81$ [0.70, 0.91], t = 15.25, $p < .0001$). The standardized coefficient was 0.79 [0.68, 0.89].

9.4. Brief discussion

As predicted by H1, "Think of many ways" increased difficulty-in-generation (relative to "Think of few"). The other hypotheses were, *on the face of it*, not supported. That is to say, statistical tests of interaction provided little evidence to suggest that the effect of difficulty-in-generation and indirect effect of condition depended on baseline PGA. However, when baseline PGA was controlled, difficulty-in-generation was negatively associated with posttest PGA (albeit not quite to a statistically significant degree). In addition, as predicted by the inconsistent mediation model, the "Think of many" condition was estimated to have a negative indirect but a positive direct effect. The positive direct "effect" was vanishingly small. The negative indirect effect was estimated to be larger albeit within a confidence interval including zero. The partially standardised estimated effect size suggests that if this negative indirect "effect" was real, "Think of many" lowered posttest PGA by approximately one-tenth of a standard deviation by increasing difficulty-in-generation. This should almost certainly be considered a small effect (at any rate smaller than the effects in Experiments 1 and 2).

Research suggests that (i) some judgements are more malleable than others, and (ii) the more malleable a judgement is the more it may be influenced by DIG (Greifeneder et al., 2010). It is possible that PGA for a Maths-related goal is not as malleable as PGA for "doing well in end-of-year exams" (Experiment 1) or "mastering foreign language vocabulary" (Experiment 2). However, when the results of all three experiments are compared, another explanation suggests itself that is fully consistent with the hypotheses. Inspection of Table 1 and Table 3 reveals that baseline PGA was considerably higher in Experiment 3 than in Experiment 1. Indeed, the mean difference was 1.67 points (on the 0–10 scale). Perceived self-efficacy (PSE) and academic self-belief are known to decline as students progress through middle school (e.g. Schunk & Meece, 2006). This decline has been observed (for PSE) in various academic domains, including foreign languages and mathematics (e.g. Fryer & Oga-Baldwin, 2017). Anxiety and lack of confidence regarding examinations also appears to increase with age (Howard, 2020). The very first year of secondary (or middle) school may therefore represent a relative *high*-point in terms of PSE and PGA for foreign languages, maths and exams. Experiments 1 and 2 suggested that the effects of difficulty-in-generation (and "Think of many") on PGA are weaker or non-existent in students with high PGA to begin with. The (relatively) high baseline PGA of participants in Experiment 3 may therefore explain why direct and indirect effects were so small. In short, high baseline PGA may have made these students immune to the effects of DIG and "Think of Many." In this sense, the results of Experiment 3 actually *support* the hypotheses: the negative indirect effect of DIG on PGA (and positive direct effect of "Think of many") are less pronounced (or even non-existent) when students' baseline PGA is high.

10. General discussion

In the present study female students in three experiments were asked to generate (multiple) means of goal attainment. One of the primary aims was to investigate the extent to which students' perceived goal attainability (PGA) is influenced by the "availability" of goal attainment means. Previous research with undergraduates suggested that asking students to think of many (rather than few) ways to attain a goal has a negative total effect on PGA (Chang, 2010; Sanna & Schwarz, 2004; Vaughn, 1999). However, that research did not directly measure the effect of difficulty-in-generation (DIG) on PGA, let alone conduct a mediation analysis. Moreover, the aforementioned studies involved only undergraduates and shed little or no light on younger (female) students. Finally, none of those studies examined a potentially crucial individual difference – students' baseline PGA. As a result, a great deal is unknown about the effects of DIG on PGA in the context of generating means of goal attainment. The present study addressed all of the aforementioned limitations and makes several important contributions. These are discussed under the headings below.

10.1. The effect of "Availability" (of Means) on female Students' PGA

The present study provides the first experimental evidence that young female students are influenced by "availability" when generating means of goal attainment. The importance of this finding may be appreciated when one recalls how frequently children are asked to generate means of goal attainment (e.g. Beghetto et al., 2015; Brusseau et al., 2020; Conklin, 2012). The present study indicates that female students in the 12–15 age range rely not only on the number of means that they generate but also on the ease or difficulty that they have in generation. Research suggests that children begin secondary school with relatively high confidence in several academic domains (e.g. foreign languages and mathematics), which then declines in the following years (Fryer & Oga-Baldwin, 2017). The decline in perceived self-efficacy for mathematics appears to be particularly strong in females (e.g. Watt, 2004). The present study suggests that "availability" influences female students' PGA only when baseline PGA is low. If so, it is possible that female students do not begin to rely on "availability" when generating means of goal attainment until later in their school career (e.g. the second year of middle school) when they have begun to doubt their abilities.

10.2. "Think of Many," Ease-of-Retrieval, and inconsistent mediation

Another important contribution of this study relates to inconsistent mediation. Previous research has been predicated on the assumption that individuals are influenced by ease-of-retrieval in some contexts and by "numerosity" (of examples, means, etc.) in others (e.g. Schwarz et al., 1991; Tan & Agnew, 2016). However, this "either…or" perspective overlooks the fact that students are often no doubt *simultaneously* (and perhaps equally) influenced by both, albeit in opposing directions. Evidence for this inconsistent mediation hypothesis was obtained in all three studies (moderated by baseline PGA in experiments 1 and 2). This finding has important implications. First, it suggests that researchers interested in understanding ease-of-retrieval and "Think of many" should not focus on "total effects" alone. Instead they should decompose total effects into (simultaneous) direct and indirect effects. This not only makes it possible to estimate the *size* of the (indirect) effect of ease-of-retrieval but also helps to illuminate the extent to which it is countervailed by other influences (e.g. the number of goal attainment means). In the present case, it seems that the size of the effect of DIG on PGA is fairly small but increases as baseline PGA declines.

Perhaps most importantly for education, the inconsistent mediation model suggests that asking (female) students to generate multiple means of goal attainment may *not* have the intended (positive) effect on PGA. Whilst some students may experience a positive (direct) effect of generating multiple means, any such benefit may be undermined by the difficulty that they had in generation. Education professionals should therefore be aware that the metacognitive experience of difficulty-in-generation may prevent any overall gain from "Think of many". Thus, asking students to list 5 ways to do well in exams, for example, may *not* raise overall confidence for exams. For many educators this will be a counterintuitive but important revelation.

10.3. The moderating effects of baseline PGA

The third important contribution of the present study relates to baseline PGA. In Experiments 1 and 2, statistical tests of interaction suggested that the effect of difficulty-in-generation on posttest PGA (like the indirect effect of "Think of Many") was moderated by baseline PGA. It is important to note that the interactions cannot be attributed to ceiling effects. In fact, across all three experiments, 95% of students reported baseline PGA scores lower than or equal to 8.5, which means that the vast majority of students *could* have scored at least 1.5 points higher on the 0–10 PGA scale. And yet it appears that posttest PGA was affected by DIG only in students who had low PGA to begin with. This too has both practical and theoretical implications. On a practical level, when students are asked to generate means of goal attainment, special attention should be paid to those who (initially) have low expectations. Female students report greater self-doubt and lower expectations in several academic domains, e.g. mathematics and social sciences (e.g. Ringeisen et al., 2016). Special attention should perhaps therefore be paid to (older) female students, who may have relatively low baseline PGA. If the negative indirect effect could somehow be "blocked" for these students, then "Think of many" may have the desired result, i.e. higher levels of (posttest) perceived goal attainability. Various methods are available for discrediting the diagnosticity of ease-of-retrieval. For example, if those with low baseline PGA can be led to attribute DIG to some aspect of the situation (rather than to a genuine lack of means), then it may not have a negative effect on their judgements (Ruder & Bless, 2003). Future research should examine whether effects of DIG on PGA (in students with low baseline PGA) can be forestalled by discrediting DIG's informational value (e.g. "Struggling to think of strategies is perfectly normal and does *not* mean that are not many strategies).

It is important to note that the positive (direct) effect of generating many means of attainment was also apparently restricted to students low in baseline PGA. That is to say, the only students whose PGA was affected at all (either positively or negatively) appeared to be those with relatively low baseline PGA. This too is an important finding and can be related to established social psychological theories. For example, according to Brockner's (1988) "plasticity" hypothesis, individuals low in self-esteem are more susceptible to environmental cues than those high in self-esteem. Shrauger and Rosenberg (1970) found, for example, that failure considerably lowered self-esteem in undergraduate students who were low in it to begin with but had a far less damaging effect in those who were initially high. Similarly, Karl et al. (1993) found that self-efficacy judgements of students low in perceived self-efficacy (PSE) were more likely to be affected by feedback than the judgements of those initially high in PSE. The present study suggests that the "plasticity" hypothesis may apply in the context of "Think of (many)" and PGA. That is to say, female students' PGA is more likely to be affected by "Think of Many" and "availability" in those low in PGA than in those initially high.

There are several ways to explain the greater "plasticity" of low PGA. Consider the greater susceptibility to difficulty-in-generation in students with low baseline PGA. What might explain this phenomenon? One possibility is differential attributions. Individuals with low perceived self-efficacy are prone to attribute setbacks to internal or stable causes, whereas those with high PSE attribute them to external or temporary factors (Alden. 1986; Silver et al., 1995). Perceived self-efficacy and perceived goal attainability are closely related (Klein et al., 2013). Students with low PGA are therefore likely to be those with low PSE. When these students experience difficulty in generating means of goal attainment they may attribute that difficulty to an internal or stable cause, e.g. a genuine lack of means. They may then consider themselves even *less* likely to attain the goal – what Weiner (1976) termed a "low expectancy cycle." On the other hand, students high in PGA may attribute difficulty-in-generation to an aspect of the situation (e.g. the request for so many means). Attributing DIG to such a cause may leave PGA unaffected. Thus, whilst both high and low "PGA-ers" experience DIG, only the latter interpret it in such a way that it lowers PGA. Future studies could test this hypothesis by measuring students' DIG attributions.

10.4. Limitations

In the present study the goal (as well as students' age) varied from experiment to experiment primarily for the sake of generalisability. In one sense, however, this is a limitation since it means that a goal-related explanation for differences in results cannot be entirely ruled out. Future research could build on the present study by holding constant either the goal or the age of participants. Doing so would make it easier to determine whether "availability" indeed becomes more influential with age. The present study also focused on *female* students. Given concerns about low PGA in females specifically, we considered this focus to be justified. However, it is also in one sense a limitation. Research suggests that the factors affecting perceived self-efficacy in females may be different from those affecting it in males (e.g. Butz & Usher, 2015). Male students may therefore respond differently to DIG and "Think of Many." Future research should investigate this possibility. In addition, future research could build on the present study by replicating the experiments in different types of school, e.g. large state schools attended by both male and female students from different socio-economic backgrounds.

Finally, researchers may wish to include (additional) covariates in their models. In the present study, covariates (other than baseline PGA) were not measured in the experiments because we wanted to miminise attrition. Research indicates that the longer an online survey the more likely participants are to avoid beginning or completing it (e.g. Galesic & Bosnjak, 2009). Moreover, attrition rates in online surveys appear to be especially high in younger participants (Spennemann, 2022). By asking children to answer only a few questions, we were able to achieve completion rates close to 100% (98.5% across the experiments). Nevertheless, the inclusion of covariates (related to posttest PGA) should result in an even better-fitting model, thereby increasing statistical power. Researchers may therefore consider including other variables

that explain posttest PGA. For example, PGA for language-related goals may vary as a function of the focal language. Future studies might then control for "focal language" by including it as a covariate.

10.5. Conclusion

Maintaining and enhancing student PGA is an essential goal for any educator. Perceived goal attainability affects not only students' goal commitment, goal pursuit, academic interest and grades (e.g. Abdulla & Woods, 2021a; Senko & Hulleman, 2013), but also students' psychological wellbeing (e.g. Brunstein, 1993). Many educators apparently assume that asking students to generate (multiple) means of goal attainment will have a positive impact on PGA. However, previous research involving undergraduates suggests that difficulty-in-generation may play a role. The present study extends previous research by focusing on female students in a middle and high school context. The findings have important implications for students and educators alike. Those who ask (female) students to generate multiple means of goal attainment should realise that doing so may have both positive and negative effects on perceived goal attainability (at least when students have low PGA to begin with). Admittedly, students with high baseline PGA may not draw negative conclusions from difficulty-in-generation. But equally they may not benefit much from generating means. Educators inclined to ask "How else could you do that?" should be aware of likely effects.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abdulla, A. (2021). Asking students to recall success may not enhance their perceived self-efficacy. *Learning and Individual Differences, 87*, 1–9.

Abdulla, A., & Woods, R. (2021a). Obstacles vs. Resources - Comparing the Effects of a Problem-Focused, Solution-Focused and Combined Approach on Perceived Goal Attainability and Commitment. *International Journal of Applied Positive Psychology, 6*, 175–194.

Abdulla, A., & Woods, R. (2021b). The effects of current unsatisfactory performance and evaluative approach on improvement expectancy and commitment to improvement. *Motivation and Emotion, 45*, 159–170.

Abdulla, A. & Woods, R. (2021c): The Effect of Solution-Focused Scaling and Solution-Focused Questions on Expectancy and Commitment. *School Psychology Review*. Advance Online Publication DOI: 10.1080/2372966X.2021.1942196.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage Publications Inc.

Alden, L. (1986). Self-efficacy and causal attributions for social feedback. *Journal of Research in Personality, 20*, 460–473. https://doi.org/10.1016/0092-6566(86)90126-1

Anderman, E. M., & Midgley, C. (1997). Changes in achievement goal orientations, perceived academic competence, and grades across the transition to middle-level schools. *Contemporary Educational Psychology, 22*, 269–298. https://doi.org/10.1006/ceps.1996.0926

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von Kriegstein, K. (2020). Learning foreign language vocabulary with gestures and pictures enhances vocabulary memory for several months post-learning in eight-year-old school children. *Educational Psychology Review 32,* 815-850. doi:10.1007/s10648-020-09527-z.

Atherton, M. (2015). Measuring confidence levels of male and female students in open access enabling courses. *Issues in Educational Research, 25*, 81–98.

Baron, J., Granato, L., Spranca, M., & Teubal, E. (1993). Decision making biases in children and early adolescents: Exploratory studies. *Merrill Palmer Quarterly, 39*, 23–47.

Beghetto, R. A., Kaufman, J. C., & Baer, J. (2015). *Teaching for creativity in the common core classroom*. Teachers College Press.

Bong, M. (2001). Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals. *Journal of Educational Psychology, 93*, 23–34. https://doi.org/10.1037/0022-0663.93.1.23

British Psychological Society. (2014). BPS Code of Human Research Ethics (2nd ed.).

Britt, T. D., & Jex, S. M. (2015). *Thriving under stress: Harnessing the demands of the modern workplace*. Oxford University Press.

Brockner, J. (1988). *Self-esteem at work: Research, theory and practice*. Lexington books.

Brophy, J. E. (2010). *Motivating students to learn*. Routledge.

Brunstein, J. C. (1993). Personal goals and subjective well-being: A longitudinal study. *Journal of Personality and Social Psychology, 65*, 1061–1070.

Brusseau, T. A., Erwin, H., Darst, P. W., & Pangrazi, R. P. (2020). *Dynamic Physical Education for Secondary School Students (9th ed.)*. Human Kinetics: Champaign, IL.

Butz, A. R., & Usher, E. L. (2015). Salient sources of early adolescents' self-efficacy in two domains. *Contemporary Educational Psychology, 42*, 49–61.

Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self- efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology, 100,* 525–534. https://doi.org/10.1037/0022-0663.100.3.525.

Chang, C. (2010). The effects of retrieval ease on health issue judgments: Implications for campaign strategies. *Health Communication, 25*, 670–680. https://doi.org/10.1080/10410236.2010.521907

Cianca, S. (2020). *Teaching Elementary STEM Education*. Routledge.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Lawrence Erlbaum Associates.

Conklin, W. (2012). *Strategies for Developing Higher-Order Thinking Skills*. Shell Educational Publishing.

Darlington, R. B., & Hayes, A. F. (2017). *Regression analysis and linear models: Concepts, applications, and implementation*. Guilford Press.

de Groot, A. M. B., & van Hell, J. G. (2005). *The Learning of Foreign Language Vocabulary*. In *Handbook of bilingualism*: Psycholinguistic approaches (eds. J. F. Kroll & A. M. B. de Groot) (p. 9–29). Oxford University Press.

Evans, D., & Brown, J. (2009). *How to Succeed at Medical School*. Wiley-Blackwell.

Falco, L. D. (2019). An intervention to support mathematics self-efficacy in middle school. *Middle School Journal, 50*, 28–44. https://doiorg.ezproxy.rgu.ac.uk/10.1080/00940771.2019.1576580.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

Fryer, L. K., & Oga-Baldwin, W. L. Q. (2017). One more reason to learn a new language: Testing academic self-efficacy transfer at junior high school. *Frontline Learning Research, 5*, 61–75.

Fuller, E. W., McIntyre, R. B., & Oberleitner, D. E. (2013). Engineering academic performance with selective retrieval: The benefits of implied ability. *Journal of Applied Social Psychology, 43*, 2195–2202. https://doi.org/10.1111/jasp.12170

Furlan, S., Agnoli, F., & Reyna, V. F. (2013). Children's competence or adults' incompetence: Different developmental trajectories in different tasks. *Developmental Psychology, 49*, 1466–1480. https://doi.org/10.1037/a0030509

Gamble, B., Tippett, L. J., Moreau, D., & Addis, D. R. (2020). The futures we want: How goal-directed imagination relates to mental health. [PDF file]. Retrieved from: https://bc8c6d88-a2e4-417e-a6cc72d924b295cf.filesusr.com/ugd/aaa34a_1ea4acc76c30481eb366a961c778783f.pdf.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*, 349–360.

Geurten, M., Willems, S., Germain, S., & Meulemans, T. (2015). Less is more: The availability heuristic in early childhood. *British Journal of Developmental Psychology, 33*, 405–410. https://doi.org/10.1111/bjdp.12114

Gibbons, M. (2002). *The self-directed learning handbook: Challenging adolescent students to excel*. Jossey-Bass.

Gist, M. E., & Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *The Academy of Management Review, 17*, 183–211. https://doi.org/10.2307/258770

Graham, S. J. (2004). Giving up on Modern Foreign Languages? Students' Perceptions of Learning French. *Modern Language Journal, 88*, 171–191. https://doi.org/10.1111/j.0026-7902.2004.00224.x

Graham, S., Courtney, L., Tonkyn, A., & Marinis, T. (2016). Motivational trajectories for early language learning across the primary-secondary school transition. *British Educational Research Journal, 42*, 682–702.

Greifeneder, R., Bless, H., & Pham, M. T. (2010). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review, 15*, 107–141. https://doi.org/10.1177/1088868310367640

Gualtieri, S. & Denison, S. (2016). The development of heuristics in children: Base-rate neglect and representativeness. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (eds. Papafragou, A., Grodner, D., Mirman, D.,& Trueswell, J.C.). Austin, TX: Cognitive Science Society. (pp. 2117-2122).

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis* ((2nd ed.).). Guilford Press.

Hermann, A. D., Leonardelli, G. J., & Arkin, R. M. (2002). Self-doubt and self-esteem: A threat from within. *Personality and Social Psychology Bulletin, 28*, 395–408.

Howard, E. (2020). *A review of the literature concerning anxiety for educational assessments*. London: Ofqual.

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*, 1–35.

Huang, S.-C., & Zhang, Y. (2013). All roads lead to Rome: The impact of multiple attainment means on motivation. *Journal of Personality and Social Psychology, 104*, 236–248. https://doi.org/10.1037/a0031039

Jacobs, J. E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development, 62*, 166–178.

Jordan, R. R. (1997). *English for Academic Purposes A Guide and Resource Book for Teachers*. Cambridge University Press.

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data Analysis: A Model Comparison Approach to Regression, ANOVA, and Beyond* (3rd ed.). Routledge.

Karl, K. A., O'Leary-Kelly, A. M., & Martocchio, J. J. (1993). The impact of feedback and self-efficacy on performance in training. *Journal of Organizational Behavior, 14*, 379–394. https://doi.org/10.1002/job.4030140409

Kernis, M. H., Brockner, J., & Frankel, B. S. (1989). Self-esteem and reactions to failure: The mediating role of overgeneralization. *Journal of Personality and Social Psychology, 57*, 707–714.

Klein, H. J., Cooper, J. T., & Monahan, C. A. (2013). Goal commitment. In E. A. Locke, & G. P. Latham (Eds.), *New developments in goal setting and task performance* (pp. 65–89). New York: Routledge.

Kruglanski, A. W., Pierro, A., & Sheveland, A. (2011). How many roads lead to Rome? Equifinality set-size and commitment to goals and means. *European Journal of Social Psychology, 41*, 344–352. https://doi.org/10.1002/ejsp.780

Lagattuta, K. H., & Sayfan, L. (2013). Not all past events are equal: Biased attention and emerging heuristics in children's past-to-future forecasting. *Child Development, 84*, 2094–2111. https://doi.org/10.1111/cdev.12082

Lapp, D., Flood, J., & Farnan, N. (2011). *Content area reading and learning.* Lawrence Erlbaum Associates.

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin, 107*, 385–393. https://doi.org/10.1037/0033-2909.107.3.385

MacCallum, R. C., & Mar, C. M. (1995). Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin, 118*, 405–421. https://doi.org/10.1037/0033-2909.118.3.405

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593–614. https://doi.org/10.1146/annurev.psych.58.110405.085542

Massey, E. K., Gebhardt, W. A., & Garnefski, N. (2009). Self-generated goals and goal process appraisals: Relationships with sociodemographic factors and well-being. *Journal of Adolescence, 32*, 501–518.

Mazza, J. J., Dexter-Mazza, E. T., Miller, A. L., Rathus, J. H., & Murphy, H. E. (2016). *DBT Skills in Schools: Skills Training for Emotional Problem Solving for Adolescents (DBT STEPS-A).* The Guilford Press.

Molway, L., & Mutton, T. (2019). Changing mindsets in the modern foreign languages classroom: An intervention combining intelligence theories and reading strategies. *The Language Learning Journal, 48*, 598–612. https://doi-org.ezproxy.rgu.ac.uk/10.1080/09571736.2018.1554693.

Mozahem, N. A., Boulad, F. M., & Ghanem, C. M. (2020). Secondary school students and self-efficacy in mathematics: Gender and age differences. *International Journal of School & Educational Psychology, 9*, S142–S152. https://doi.org/10.1080/21683603.2020.1763877

Norrish, J. (2015). *Positive education: The Geelong Grammar School journey*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198702580.001.0001.

Replication Index, (2019). Loss of confidence in the Ease-of-Retrieval Paradigm. Retrieved from https://replicationindex.com/2019/01/03/ease-of-retrieval/.

Ringeisen, T., Raufelder, D., Schnell, K., & Rohrmann, S. (2016). Validating the proposed structure of the relationships among test anxiety and its predictors based on control-value theory: Evidence for gender-specific patterns. *Educational Psychology, 36*, 1826–1844.

Ruder, M., & Bless, H. (2003). Mood and the reliance on the ease of retrieval heuristic. *Journal of Personality and Social Psychology, 85*, 20–32. https://doi.org/10.1037/0022-3514.85.1.20

Sanna, L. J., & Schwarz, N. (2004). Integrating Temporal Biases: The Interplay of Focal Thoughts and Accessibility Experiences. *Psychological Science, 15*, 474–481. https://doi.org/10.1111/j.0956-7976.2004.00704.x

Schunk, D. H., & Meece, J. L. (2006). Self-efficacy development in adolescence. In *Self-efficacy beliefs of adolescents* (eds. F. Pajares & T. Urdan) (pp. 71–96). Information Age Publishing.

Senko, C., & Hulleman, C. S. (2013). The role of goal attainment expectancies in achievement goal pursuit. *Journal of Educational Psychology, 105*(2), 504–521. https://doi.org/10.1037/a0031136

Paratore, J. R., & McCormack, R. L. (2005). *Teaching literacy in second grade.* Guildford Press.

Parks, A. N. (2015). *Exploring mathematics through play in the early childhood classroom.* Teachers College Press.

Putwain, D., & Daly, A. L. (2014). Test anxiety prevalence and gender differences in a sample of English secondary school students. *Educational Studies, 40*, 554–570.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*, 195–202. https://doi.org/10.1037/0022-3514.61.2.195

Shrauger, J. S., & Rosenberg, S. E. (1970). Self-esteem and the effects of success and failure feedback on performance. *Journal of Personality, 38*, 404–417.

Sill, L. M., & Smith, J. (2017). *Leveled Text-Dependent Question Stems: Mathematics Problem Solving.* Shell Educational Publishing.

Silver, W. S., Mitchell, T. R., & Gist, M. E. (1995). Responses to successful and unsuccessful performance: The moderating effect of self-efficacy on the relationship between performance and attributions. *Organizational Behavior and Human Decision Processes, 62*, 286–299. https://doi.org/10.1006/obhd.1995.1051

Stormont, M., Reinke, W. M., Herman, K. C., & Lembke, E. S. (2012). *Academic and behaviour supports for at-risk students: Tier 2 interventions.* Guildford Press.

Sung, Y.-T., Chao, T.-Y., & Tseng, F.-L. (2016). Reexamining the relationship between test anxiety and learning achievement: An individual-differences perspective. *Contemporary Educational Psychology, 46*, 241–252.

Tan, K., & Agnew, C. R. (2016). Ease of retrieval effects on relationship commitment: The role of future plans. *Personality and Social Psychology Bulletin, 42*, 161–171. https://doi.org/10.1177/0146167215617201

Tormala, Z. L., Petty, R. E., & Briñol, P. (2002). Ease of retrieval effects in persuasion: A self-validation analysis. *Personality and Social Psychology Bulletin, 28*, 1700–1712. https://doi.org/10.1177/014616702237651

Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A New Approach to Assessing Strategic Learning: The Case of Self-Regulation in Vocabulary Acquisition. *Applied Linguistics, 27*, 78–102. https://doi.org/10.1093/applin/ami046

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

UK Government, (n.d.). The national curriculum. Retrieved from https://www.gov.uk/national-curriculum/key-stage-3-and-4.

Vaughn, L. A. (1999). Effects of uncertainty on use of the availability of heuristic for self-efficacy judgments. *European Journal of Social Psychology, 29*, 407–410. https://doi.org/10.1002/(SICI)1099-0992(199903/05)29:2/3<407::AID- EJSP943>3.0.CO;2-3.

Watt, H. M. G. (2004). Development of adolescents' self-perceptions, values, and task perceptions according to gender and domain in 7th- through 11th-grade Australian students. *Child Development, 75*, 1556–1574.

Watt, H. M. G. (2006). The Role of Motivation in Gendered Educational and Occupational Trajectories Related to Maths. *Educational Research and Evaluation, 12*, 305–322.

Weiner, B. (1976). An attributional approach for educational psychology. *Review of Research in Education, 4*, 179–209.

Weingarten, E., & Hutchinson, J. W. (2018). Does ease mediate the ease-of-retrieval effect? A meta-analysis. *Psychological Bulletin, 144*, 227–283. https://doi.org/10.1037/bul0000122

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*, 552–565. https://doi.org/10.1037/0012-1649.27.4.552

Wyra, M., & Lawson, M. J. (2018). Foreign language vocabulary learning using the keyword method: Strategy and meta-strategy knowledge. *The Language Learning Journal, 46*, 605–621. https://doi.org/10.1080/09571736.2018.1503138

Yeager, D. S., Krosnick, J. A., Visser, P. S., Holbrook, A. L., & Tahk, A. M. (2019). Moderation of classic social psychological effects by demographics in the U.S. adult population: New opportunities for theoretical advancement. *Journal of Personality and Social Psychology, 117*, e84–e99. https://doi.org/10.1037/pspa0000171

Spennemann, D. H. R. (2022). Persistence and attrition among participants in a multi-page online survey recruited via Reddit's social media network. Social Sciences, 11 (2), [31].