Contemporary Educational Psychology 66 (2021) 101993

Contents lists available at ScienceDirect

Contemporary Educational Psychology

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

Process matters: Teachers benefit their classrooms and students when they deliver an identity-based motivation intervention with fidelity☆

Daphna Oyserman [a,*], S. Casey O'Donnell [a], Nicholas Sorensen [b], Kimberly M. Wingert [c]

[a] University of Southern California, USA
[b] Summitlab and American Institutes for Research, USA
[c] Consortium of Universities of the Washington Metropolitan Area, USA

ARTICLE INFO

Keywords:
Motivation
Self-regulation
Academic risk
Intervention
Fidelity
Social-emotional competencies
Train-the-trainer

ABSTRACT

Students value school success but often experience classroom norms implying that learning is easy and succeeding in school is not difficult. Applying an identity-based motivation (IBM) lens highlights three ways succeed-with-ease-not-effort norms can undermine students' grades and increase their risk of course failure. Succeed-with-ease-not-effort norms reduce the likelihood that students experience school as relevant to their future goals, experience right now as the time to get going, and difficulties as signals of schoolwork's importance, not its impossibility. To support student academic outcomes, we examine *Pathways-to-Success*, a classroom-level intervention operationalizing IBM theory in a 3-cycle, 3-year development design (N = 1142 8th-graders, 87% low-income families, 64% Latinx, 20% African American). We document that *Pathways-to-Success* can be sustainable; our middle school teachers implemented and taught other teachers to implement *Pathways-to-Success*. We use structural equation models to show that effects are due to the theorized process; teachers who implemented with more signal clarity supported academic success by bolstering their students' identity-based motivation. We operationalized signal clarity as a mean of five fidelity components (dosage, adherence, quality, responsiveness, receipt). Signal clarity matters; students experiencing *Pathways-to-Success* with a clearer signal have a higher identity-based motivation score. Higher identity-based motivation yields better school outcomes.

1. Introduction

People who have more education are better off. They are happier, healthier, and more likely to be employed and financially secure than people who do not finish high school or college (Brunello, Fort, Schneeweis, & Winter-Ebmer, 2016; Conti, Heckman, & Urzua, 2010; Duncan & Murnane, 2011). Students seem to know this; most expect to earn a college degree (ACT, 2018; Cowan, 2018; Jacob & Linkow, 2011; McFarland, Hussar, Wang, Zhang, Wang, Rathbun, & Ossolinski, 2018; Wolniak, Davis, Williams, & Casano, 2016; for a review, Oyserman, 2013). These school-focused possible identities matter: students who expect that school will be central to the person they will become are at lower risk of poor grades and course failure (Oyserman, Bybee, Terry, & Hart-Johnson, 2004; Horowitz, Oyserman, Dehghani, & Sorensen, 2020) and complete more schooling (Beal & Crockett, 2010). But school-focused possible identities are not enough. Contexts differ in how much they support students experiencing their school-focused identities as relevant, current action as necessary, and difficulty as implying that succeeding at school tasks is important (not impossible). One such context is the classroom. The norm students may infer from classroom interactions is that school success comes with ease and without much effort (Xu, 2006, 2007). This succeed-with-ease-not-effort norm mismatches the reality that deep learning takes time and entails difficulty (Soderstrom & Bjork, 2015), and that difficulty can be a signal of value (Fisher & Oyserman, 2017; O'Donnell, Yan, Bi, & Oyserman, 2021). Applying an identity-based motivation lens highlights three ways this

☆ *Note*: We thank Nick Yoder for directing fidelity coding and Eric Horowitz for the syntax to compute fidelity, the involved schools, teachers, and students, and especially our teacher-trainers: Lahnna Addington, Cathy Cunningham-Yee, Sophia Panos, Joshua Silverman, and Jennifer Thompson. The research reported in this paper was funded by the U.S. Department of Education, Institute of Education Sciences, grant number # R305A140281, to Oyserman and Sorensen. Declarations of interest: none. This paper has not been submitted elsewhere for consideration for publication.

* Corresponding author at: University of Southern California, Mind and Society Center, 635 Downey Way, Verna & Peter Dauterive Hall, Suite 205, Los Angeles, CA 90089-3333, USA.
E-mail address: oyserman@usc.edu (D. Oyserman).

https://doi.org/10.1016/j.cedpsych.2021.101993

Available online 1 July 2021

mismatch undermines students (Aelenei, Lewis, & Oyserman, 2017; Oyserman et al., 2017). First, it undermines a student's ability to notice that now is the time to get going (Oyserman, 2009). Second, it undermines students' interpretation of their experiences of difficulty, setting them up to endorse difficulty-as-impossibility rather than consider a difficulty-as-importance perspective (Oyserman, 2007). Third, given the social nature of school-relevant identities (Oyserman, 2007), it triggers a negative spiral in which students self-handicap (Clarke & MacCann, 2016; Hirt & McCrea, 2009; Jones & Berglas, 1978) and disrupt one another (Duncan & Murnane, 2011; Rowan, 2011).

In the current paper, we build on identity-based motivation theory and use a 3-cycle development design to test a way to mitigate this negative spiral by changing the classroom context that fosters it. We document that teachers can successfully teach other teachers to implement a short, whole-classroom identity-based motivation intervention, and that investing in implementation matters. Implementing better provides a clearer signal. Clearer signals yield more effect on the theorized process (identity-based motivation) and outcomes (better grades and reduced likelihood of course failure).

2. Identity-based motivation

Identity-based motivation (IBM) theory focuses on the situated sensitivity and motivational power of the self (Oyserman, 2007, 2015; Oyserman, Elmore, & Smith, 2012). People are motivated to make sense of their experiences and act in ways that fit their temporal sense of their personal and social identities. Situations shape what people infer about themselves (how "we" act, and what "we" believe). In the case of classrooms, students should be sensitive to cues about the centrality and relevance of their school-focused possible identities to the current situation and how to make sense of their metacognitive experiences. They use these cues as implicit answers to their questions about whether what happens at school is relevant to their current and future identities and how to make sense of their experiences of difficulty.

2.1. Evidence to date

Several experiments tested these IBM assertions about the motivational force and context-sensitivity of identities. In these experiments, researchers randomly assigned students to one of two groups. One group of students received cues meant to lead them to infer that school is identity-relevant. The other group received cues meant to lead them to infer that school is identity-irrelevant. Sometimes the identities were future-focused (e.g., "college me", "adult me", Oyserman, Destin, & Novin, 2015; Oyserman et al., 2018; Nurra & Oyserman, 2018). Other times, the identities were social (e.g., race-ethnicity, social class, gender, (Oyserman, Bybee, Terry, & Hart-Johnson, 2004; Oyserman et al., 2012; Oyserman, Gant, & Ager, 1995; Oyserman, 2009). Students led to infer that school is irrelevant to their future identities were more likely to agree that facing difficulty in a school task implies that it is impossible to succeed (Oyserman et al., 2015). They were less engaged with schoolwork and attained worse grades (Nurra & Oyserman, 2018). Students who were led to infer that school is relevant to their social identities engaged more with their schoolwork (Destin and Oyserman, 2009; Elmore & Oyserman, 2012; Oyserman et al., 1995). For example, middle school boys were more likely to spend time on schoolwork and experience school as central to their futures after they saw graphs showing earnings by gender (men earn more) instead of one showing educational attainment by gender (women attain more). The implication is that the nature of the graphs shaped the inferences boys made as to whether schoolwork is a 'boy' thing to do (Elmore & Oyserman, 2012).

Several other experiments tested the IBM assertions about the effect of context on the likelihood that students will infer that school is the path and they need to get going and invest immediately in their future selves. In these experiments, college students saw descriptions of the world as certain or uncertain and were asked to consider their certainty or uncertainty about themselves (Smith, James, Varnum, & Oyserman, 2014). Students led to feel self-uncertain did not take action to work on their school-focused possible identities. Neither did students led to feel certainty about the world. In contrast, students who were led to feel self-certain in the context of an uncertain world described school-focused identities, planned to block out more time for studying, and used a resume-builder rather than play a computer game. The inference they seemed to make is that they had better take immediate action to work toward their goals, given that they had the skills, and that the world was an uncertain place.

Several experiments tested the IBM predictions about the consequences of students' norm sensitivity (Aelenei, Lewis, & Oyserman, 2017; Smith & Oyserman, 2015). For example, Smith and Oyserman (2015) had students report the frequency with which they interpreted their difficulty working on a school task as signaling importance, that what they were working on was worth their effort because it was important to them. Before doing some, half of the students were randomly assigned to use a high-frequency scale (ranging from less than ten times a month to 31 or more times a month). The other half of students were assigned to use a low-frequency scale (ranging from once or twice a month to eleven or more times a month). Researchers did not say that the frequency scale represented how typical it was for others to think this way. But students seemed to draw this inference, inferring their relative standing given where they were on the frequency scale they saw. The same response could imply higher-standing (on a low-frequency response scale) or lower-standing (on a high-frequency response scale). Students used this information to infer whether they believed that difficulty meant importance. In the high-frequency scale condition, they drew the inference that they believed difficulty-means-importance less frequently than others, and their self-rated difficulty-as-importance score declined. This inference affected their sense that investing in school is central to their identity and the actual time they spent on subsequent difficult tasks.

Over time, the effects of these single-cue experiments have on school engagement and attainment fade (thus Nurra & Oyserman, 2018, show effects fade within months of the cue). In contrast, we found evidence that effects on well-being and attainment become longer-lasting when each of the core elements of identity-based motivation is cued repeatedly as part of a manualized series of 12 group-based activities in classrooms (Oyserman, 2015; Oyserman, Bybee, & Terry, 2006; Oyserman, Terry & Bybee, 2002). Table 1 summarizes the activities and take-home points in each session (the full manualized intervention is available in Oyserman, 2015). Fig. 1 provides a session-level depiction of the activities in this brief classroom intervention called *Pathways-to-Success* (*Pathways*), color-coded for each core IBM element. Activities facilitate experiencing school as identity-relevant (blue), readiness to act and use strategies to work toward future identities (now is the time, yellow), and difficulty as implying the importance, not the impossibility of success in school (red). As shown in the mix of colors, activities often involve more than one aspect of identity-based motivation. Table 2 summarizes the underlying structure of the *Pathways* activities and learning environment.

In the previous tests, pairs of researcher-trained college students (Oyserman et al., 2002), pairs of researcher-trained community members (Oyserman et al., 2006), or researcher-trained teachers working alone (Horowitz, Sorensen, Yoder, & Oyserman, 2018) implemented *Pathways* in classrooms. One test was a randomized trial (Oyserman et al., 2006). It showed that being randomly assigned to *Pathways* reduced the likelihood that student grades would drop over time and that they would be held back by changing possible identities and that effects were long-lasting, persisting into high school. A version of this intervention was used to improve outcomes among incoming undergraduates seeking to complete undergraduate degrees in business administration who were either first-generation to college, from low-income families, or attended high schools lacking advanced mathematics courses (Lewis & Yates, 2019).

Table 1
Thumbnail sketch of each *Pathways* session activities and take-home point.

Session	Classroom activity flow	Take-home point
1. Setting the Stage & Introduction of Skills and Abilities	Students are paired up and briefly interview one another on the skills or ability they each have that will help them complete the school year successfully (e.g., "well organized," "positive attitude"). Then each student introduces his or her interview partner in terms of these skills.	We all care about school, and we have a skill or ability to work on our "successful in school" possible self.
2. Adult Images	Students pick photographs that fit their adult "images"—images of what their adulthood will be like. Photographs include the four domains of adulthood lifestyle (e.g., homes), job (e.g., working at various jobs), relationships (e.g., family, friends), and community engagement (e.g., volunteering, voting). Photographs include both genders and match the racial-ethnic makeup of the school. Domains of adulthood emerge from clustering student responses and having students name these clusters.	We all have images of ourselves as adults in the far future.
3. Positive and Negative Forces	Students draw or write about positive and negative forces—people or things that energize them to work toward their possible identities by showing what to do or what not to do.	Everyone faces obstacles and difficulties; positive and negative forces help by laying out paths to take or avoid and ways to handle obstacles and examples of what not to do.
4. Timelines 1	Students list things that they expect to happen in their future and begin to organize them chronologically. They learn about forks and roadblocks and describe one of each that might occur in their futures, including ways around potential roadblocks.	The present and future are linked on a path with forks (choices) and roadblocks (obstacles). Choices set up which futures are possible, obstacles are not chosen but happen. We have to figure out ways around obstacles to make progress and create paths.
5. Timelines 2	Students use their drafts to draw their timelines into the future including a fork and an obstacle and a way around it.	The present and future are linked on a path with forks (choices) and roadblocks (obstacles). Choices set up which futures are possible, obstacles are not chosen but happen. We have to figure out ways around obstacles to make progress and create paths.
6 Possible selves and strategies	Students map out their expected and to-be-avoided possible selves and strategies for next year on a pathways board.	Strategies are actions you are taking now or could take to become your next year's possible self.
7. Pathways to the Future	Students complete pathways boards to concretize the link between current strategies for action, next year's possible selves, and adult possible selves.	Strategies I'm doing (or could be doing) now to get to my next year's possible self also help me get to my adult possible self.
8. Action Goals	Students write action goals, linking next year and adult possible selves with actions they can take right away in a specific time and place to concretize the plan. They do this using an easy-to-recall formula (because… I will… when…).	We have some control over possible selves, but not our hopes and dreams. We gain some control by linking the future with the present through specific action paths.
9. Puzzles	Students learn about the idea of inoculation from experiencing difficulty as meaning that something is impossible or not for you. They break down problems that seem impossible and use strategies to solve them.	Difficult things can seem impossible, not worth your time; but having difficulty can be a signal of importance. Simple strategies like breaking down into parts can help.
10. Solving Everyday Problems	Students practice being inoculated from experiencing difficulty as meaning something is impossible. They work on an everyday problem at school and write about a school problem they have. In each case, they work together using the identity-based motivation skills they have learned to consider how to break the problem down.	Everyday problems can be broken down using the skills you have to consider what is the adult possible self, what is the next year possible self, what is the positive and negative force, the choice point or obstacle, and what are strategies to get around it.
11. Everyday problems: High school and beyond	Students brainstorm what is needed to finish high school, see the requirements, brainstorm what is needed to get to college, and see the requirements.	You can identify the steps to get from 8th-grade to graduating high school.
12. Wrapping up and Moving Forward	Students order and name each *Pathways* session and what it was about. They consider what they liked and what they would improve. Activities provide a bird's eye view, closure, and reinforcement of the three main IBM ingredients.	What I do now matters for attaining my next year and adult possible selves. Possible selves that are linked to strategies and to time and place of action become action goals. There are forks (choices) and roadblocks (failures) along the way. It will be difficult and may feel impossible, but asking questions helps break down what I need to find out and helps me connect to others – positive forces and models – as well as to learn from negative forces and models of what not to do.

2.2. Next steps toward a sustainable IBM-based intervention

While promising, these earlier studies entailed researcher-trained college students or community-members or the researchers themselves delivering the intervention. These studies document that *Pathways* can yield intended effects. In this section, we take the next step and outline the evidence for reproducibility and scalability, replicability, and sustainability of *Pathways*. First, we consider available evidence regarding the reproducibility and scalability of *Pathways*. Next, we consider the evidence for effect replicability. Then we turn to questions regarding the sustainability of delivery in a school setting over time.

Regarding reproducibility, the *Pathways* intervention is manualized (Oyserman, 2015). Having a manual means that, in principle, it is possible to reproduce the *Pathways* experience. Regarding scalability, college students (Oyserman et al., 2002), community members (Oyserman et al., 2006), and teachers (Horowitz et al., 2018) successfully implemented *Pathways* after experiencing a researcher-led structured cycle of learning. So, the evidence suggests that *Pathways* is scalable.

Regarding the replicability of results, the above-described experiments tested aspects of the theorized process, showing the effects of

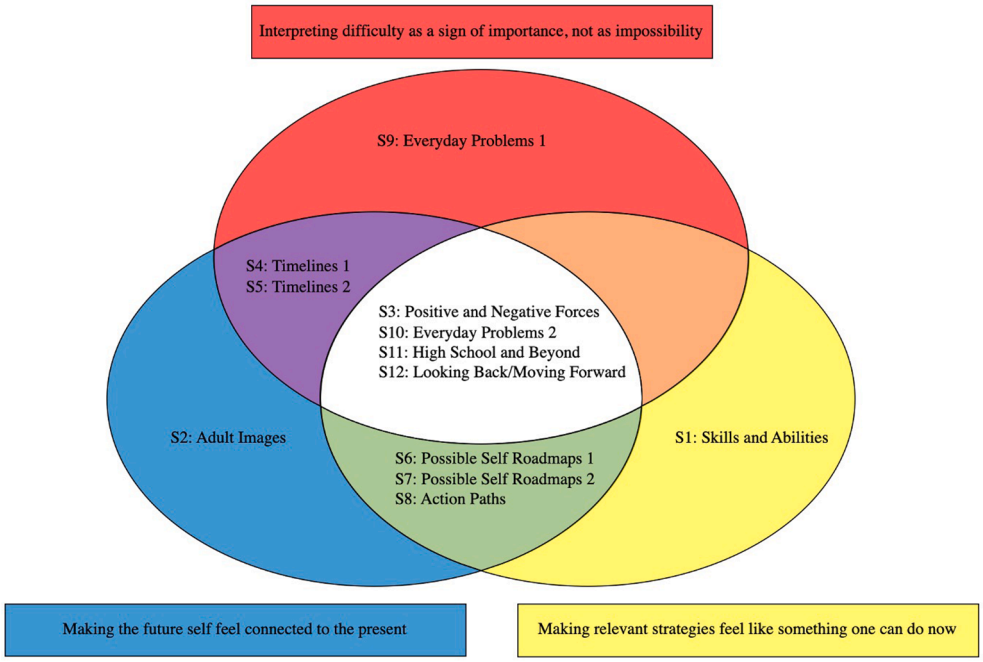

Fig. 1. *Pathways-to-Success* Sessions, Labeled S1 to S12, Color-Coded by Active Ingredient.

Table 2
The Underlying Structure of the *Pathways-to-Success* Intervention Represented as Three Columns.

Each session contains the same sequence	*Each session is…*	*IBM Constructs are…*
Welcome	Stable in its underlying structure	Introduced using student-elicited ideas.
Student-led recap of the prior session	Unique in content	Evoked from student ideas to create a joint understanding
Link to the current session	Sequenced so sessions build on one another	Connected across sessions
Introduce new terms through a student-led description of prior knowledge	Active experience-based (students do something, create something).	Operationalized in different ways across sessions so students can attain a deeper understanding.
Active experience-based learning	A mix of group and individual activities	Naturally emergent from the session activity
Student-led, teacher-scaffolded recap	Time-limited to a class period	Evoked through student discussion
Link to next session	Focused on one or more IBM constructs	Learned through experience, practice, individual and group engagement

manipulating each component of identity-based motivation separately. The randomized trial measured part of identity-based motivation, possible identities. It showed that effects were due to changes in this measured aspect (e.g., Oyserman et al., 2006). So, in principle, this means results are due to at least part of the theorized process.

Regarding sustainability, *Pathways* is brief and inexpensive. It does not require expensive ongoing resources or financial support to continue implementing with each new cohort of students. The implication is that it could be sustainable in schools with teachers who have learned how to run *Pathways*. In this paper, we consider three remaining gaps regarding sustainability, reproducibility, replicability, and scalability.

3. Moving from showing that an intervention could work to sustainability, reproducibility, replicability, and scalability

In this section, we consider three remaining gaps regarding sustainability, reproducibility, replicability, and scalability which we term staff turnover, quality assurance, and fuller test of the theorized process model.

3.1. Staff turnover

In the current researcher-trained model, when college students complete their semester-long commitment, community members find other jobs, and teachers leave the profession, shift schools, or transfer to different grades and responsibilities within the same school, they take the skills, capacities, capabilities, and competencies they gained in training with them. As new people take over, they require training. A researcher-led training model creates a resource bottleneck because the researcher is unlikely to provide such on-demand training. One way around the researcher-trained resource bottleneck is for teachers who have learned how to run *Pathways* to pass their knowledge on to other teachers (termed a train-the-trainer model).[1]

3.2. Quality assurance

We noted above that people from different walks of life (college students, community members, and teachers) have implemented *Pathways*. Examining these prior reports, we also find that people from different walks of life received different intensities of training and support and implemented *Pathways* differently. Teachers received a 2-day training with weekly video support sessions (Horowitz et al., 2018). College students (Oyserman et al., 2002) and community members (Oyserman et al., 2006) received a week of training and in-person support before delivering each session. Teachers worked alone with their entire class. College students and community members worked in pairs and implemented *Pathways* to half the classroom at a time. Signal quality

[1] Throughout, we use the term train-the-trainer to describe teachers training other teachers in the skills and abilities needed to successfully provide *Pathways* in their classrooms. Other terms, including "cascade training", or "pyramid training" are used in the fields of child welfare, psychology, policy, and implementation science literature to describe this same process (Herschell, Kolko, Baumann, & Davis, 2010; McHugh & Barlow, 2010; Weingarten et al., 2018). In the review process, a reviewer voiced concern that training is rote or mechanical. Training could entail lower-level or rote skill but is also used to describe attaining higher-level and complex skill. Our train-the-trainer methods entail the development of all the skills and capacities needed to provide Pathways with high skill as we detail in our section on fidelity.

might differ based on these factors. We did not find direct comparisons of signal quality across these variations. But large-scale variability in signal quality could undermine the extent that *Pathways* is scalable and reproducible and how replicable *Pathways* results are. Addressing this gap requires a better understanding of what signal quality entails.

3.3. Fuller test of the theorized process model

We noted above that research to date only tested some aspects of the theorized identity-based motivation process. That is, we know that students randomly assigned to treatment rather than to school-as-usual are less likely to fare poorly in school and show fewer symptoms of depression and that student possible identity scores mediated these effects. But research to date has not assessed whether *Pathways* affects how students interpret their experiences of difficulty – no measure was available at the time of the initial test (see Oyserman et al., 2006). If *Pathways* successfully operationalizes identity-based motivation theory, then students receiving a clearer signal should show a more positive response on a measure operationalizing identity-based motivation more fully. Moreover, clearer signals should support student academic success through increased identity-based motivation.

3.4. Intervention fidelity and measuring process

The formal term for signal clarity is fidelity. Within a medical model, people often reduce the scope of the term fidelity to two aspects (compliance, and adherence). That is, they focus on compliance with dosage (e.g., take this much medication) and adherence to a protocol (e.g., take this medication with liquid four times daily for 2 weeks). These aspects are valuable but insufficient markers of fidelity for program and intervention researchers. Instead, for program and intervention researchers, fidelity is better understood as the average of both the commonly considered aspects of dosage and adherence, and three less commonly considered aspects, termed quality of implementation, student responsiveness, and receipt (see, Bellg et al., 2004; Crosse, Williams, Hagen, Harmon, Ristow, DiGaetano, & Derzon, 2011; Dane & Schneider, 1998; Dusenbury, Brannigan, Falco, & Hansen, 2003; King-Sears, Walker, & Barry, 2018; Mowbray, Holter, Teague, & Bybee, 2003; O'Donnell, 2008; Resnick et al., 2005). *Dosage* is how much of the planned sessions occur in the planned time sequence. *Adherence* is the extent to which teachers implement manual-described activities and sequencing. *Quality* is how much teachers engage with clarity, use a positive tone, elicit student examples, and use students' ideas to concretize core concepts. *Responsiveness* is how much students respond as intended during activities. *Fidelity of receipt* is student certainty and confidence in core intervention activities and messages. In that sense, when teachers deliver *Pathways* with fidelity, they are providing a learning environment that students experience as collaborative. As detailed in Table 2, in *Pathways*, teachers scaffold student learning by giving students experiential opportunities, building on student examples to define core terms, linking experiences to the core constructs, and operationalizing these constructs repeatedly as each re-appears across sessions in different guises.

In some ways, fidelity scores can be used the way manipulation checks are used in experiments. They can provide some sense that effects are due to the intended process. However, this comparison has limitations. Significant manipulation checks verify that experimental manipulations changed response on a proximal construct, implying that outcomes are due to the theorized process. Fidelity does not entail a significance test and assesses both what happened and proximal change. In practice, researchers often rely on describing some aspects of fidelity as an assurance that implementation included some proportion of the planned activities (e.g., Eddy et al., 2017). A five-factor fidelity score provides more than that. It also addresses the question of whether the intervention was delivered with quality, whether participants responded as intended, and whether they experienced the immediate changes posited by the intervention.

Because of this broader scope, descriptive information about fidelity can address reproducibility. Confidence in reproducibility should be higher when fidelity is similar across differences in the people who are taught to use the intervention, in how they are taught, and in how the intervention is delivered. Fidelity scoring is also critical in addressing questions of process. Scoring fidelity and measuring process variables allows researchers to know if effects have to do with the planned intervention and work through the theorized process. Low fidelity means that the intervention was not provided as intended, something other than the planned intervention occurred. Low fidelity makes interpretation of positive and negative results problematic because it is not clear what happened. If the theory itself is documented to affect change in experiments and other non-intervention methods, then low fidelity signals that the theory is not being translated clearly or that changes in how people are taught to use the intervention or in how they implement it are needed. If positive effects are found when fidelity is lower than an agreed-upon threshold, that implies that whatever worked was outside of the theorized process model. In contrast, if positive effects are found when fidelity is at or above an agreed-upon threshold, that implies that the theorized process is at least partly responsible for effects. Although not yet studied as such, as fidelity increases beyond an agreed-upon threshold, higher fidelity should show a stronger effect because it represents a clearer signal or representation of the theorized process.

Unfortunately, intervention research typically does not report or use fidelity in inference-making (Durlak & DuPre, 2008; Durlak, Weissberg, & Pachan, 2010; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Vivalt, 2020) or measures of the theorized process (Leeuw & Vaessen, 2009; Prestwich, Webb, & Conner, 2015). Both omissions make it harder for people to iteratively learn from prior experience. Instead of measuring fidelity and assessing the effects of fidelity on process measures, researchers assume that attaining the predicted outcomes implies that the intervention was appropriately delivered, and the theory-based process was supported. That this is not the case is driven home by the high heterogeneity of results across deliveries of interventions (Vivalt, 2020). Some results are positive, some negative, and some null (middle school-based intervention effects range from -0.15 *SD* to $+0.38$ *SD,* Boulay et al., 2018). Without scored fidelity and measures of process, researchers cannot diagnose sources of heterogeneity in results across attempted replications except through post-hoc exploration (e.g., Allcott, 2015; Borman, Grigg, Rozek, Hanselman, & Dewey, 2018; Sisk, Burgoyne, Sun, Butler, & Macnamara, 2018). If fidelity is not scored or only dosage or adherence aspects are documented, heterogeneity of results could be due to differences in what people did when they delivered the intervention. Lacking this information, people can look post-hoc for explanations to attempt to describe differences in treatment effects across recipient subgroups (e.g., race-ethnicity, poverty, achievement, Sisk et al., 2018), site-level contexts (e.g., sites that volunteer, Allcott, 2015; school-level achievement-gaps by race-ethnicity, Borman et al., 2018), or who implemented the intervention (e.g., government or non-government employees, Vivalt, 2020). If theorized process measures are measured, measured and shown not to be affected by the intervention, or measured and shown to affect outcomes but not in the expected direction, the source of heterogeneity may be misspecification of the theory and hence its operationalization to intervention. But neither possibility can be tested without scored fidelity and process measures.

Lacking these building blocks, researchers cannot use heterogeneity to attain a clearer understanding of what works and why. They cannot know how to replicate positive effects or how to avoid replicating negative ones. This can be harmful to intervention recipients and wasteful of limited resources. In this sense, the field's failure to provide a basis to iteratively learn from experience through full five-factor fidelity assessment and assessment of process measures is costly in terms of the opportunity costs for human lives and in terms of the limited monetary resources for intervention that could have been better applied. Fig. 2 shows the difference between this black-box approach going directly

from knowing that an intervention occurred to outcomes (top panel) and our process approach (bottom panel) of measuring both fidelity and its effect on process measures.

3.5. How much fidelity is enough?

Rules-of-thumb exist to provide benchmarks regarding the size (Cohen, 1988) or likelihood (Gignac & Szodorai, 2016) of effects but not for fidelity. To create a working model for fidelity, we build from the conclusions Durlak and DuPre (2008) drew from their extensive *meta*-analytic reviews to suggest a plausible rule-of-thumb for fidelity effects. They found that non-researchers rarely implement with over 80% fidelity, and if implemented with under 60% fidelity, interventions are unlikely to yield the intended effects. As rules-of-thumb, we can say that scores over 80% are implausible targets, and scores under 60% raise doubts as to whether implementation yields a strong enough signal to produce the intended effects. In these cases, observed effects may well have resulted from something other than the theorized intervention. Moving from 60% toward 80% can be thought of as improving signal clarity, a fuller representation of the theorized active ingredients with fewer other ingredients. If the theorized process is occurring, this should yield more effect on outcomes.

4. Current study

We present our theorized process model in Fig. 3. We test scalability and sustainability (H1) and process (H2) predictions by scoring fidelity and attaining measures of the theorized process.

H1: When teachers implement and teach other teachers to implement *Pathways*, they can attain threshold or higher fidelity.
H2: *Pathways* operationalizes the IBM theory. The stronger the signal, the larger the effect on measures of students' identity-based motivation, and hence, their academic outcomes (and potentially, their self-regulatory competence).

4.1. Sample

Students (N = 1142, 51% girls) were 8th-graders in AY14-15, AY15-16, or AY16-17 enrolled in one of 10 high-poverty Chicago public schools, participating in one of 40 classrooms with an average class size of 28.70 (SD = 4.68) students.[2] Students were mostly from very low-income households as defined by the criteria of receiving free/reduced-price lunch (87%), racially-ethnically diverse, and, like student bodies in urban school districts, student bodies were "majority-minority" (64% Latinx; 20% Black; 16% White, Asian, Multiracial or other).[3] Teachers (N = 28) were the 8th-grade teachers who taught one or more core subjects (Math n = 9, English n = 7, social studies n = 5, science n = 10, special education n = 1)[4] in each of the participating schools. As detailed below, their students generally rated them as enthusiastic, warm, clear, and knowledgeable.

4.2. Study procedure

4.2.1. Overview of timeline and activities

Schools entered in Cycle-1 (AY14-15, 8 teachers, 2 schools), Cycle-2 (AY15-16, 8 teachers, 3 schools), or Cycle-3 (AY16-17, 12 teachers, 5 schools). Cycles entailed training each 8th-grade teacher who taught a core subject over the summer and each teacher implementing *Pathways* in their advisory classroom (with video recording) during the first weeks of the school year. An American Institutes for Research (AIR) research assistant or an outside videographer set up recording before and retrieved recording after each session for each teacher (472 videos).

We schematize our timeline of activities in Fig. 4. CPS provided school records. We collected baseline measures of identity-based motivation. Then teachers implemented *Pathways* in a single classroom period, twice a week for 6 weeks. We collected student end-of-*Pathways* surveys (October) and year-end surveys (a repeat of baseline, April/May).

4.2.2. Procedure for applying iterative development principles

We engaged in a 3-cycle iterative development process. We learned from our teachers both by hearing from them directly in end-of-year interviews and by carefully watching the video to figure out places in which the manual was unclear, or teachers struggled to actively evoke student experience and reflect student experiences to them in ways that scaffolded deeper conceptual learning. To increase teachers' comfort in sharing critical insights, interviews were led by the third author and other members of our team who did not deliver or support training and focused on eliciting critical comments so that we could improve. We outline the specific changes made at the end of each development cycle in Tables 3 and 4. Table 3 describes the training and weekly support, who was taught by whom, and how. Table 4 describes what we learned from our teachers and how we used these insights to improve the *Pathways* manual, resources, and other aspects of the structure of implementing *Pathways* and supports for teacher-educators (teacher-trainers).

After each cycle, teachers who delivered with the highest fidelity were asked to teach teachers in the next cycle (two teachers after Cyc1e 1, four after Cycle 2). All agreed. Moreover, the schools in which our teacher-educators were located elected to continue delivering *Pathways*. Each year, continuing teachers were invited to participate again in training if they wanted a refresher, but none chose to do so. In total, as summarized in Table 5, our teachers delivered *Pathways* to 40 classrooms (AY 2014–15, n = 8, AY 2015–16 n = 11, AY 2016–17 n = 21). Three teachers (one of them a trainer) delivered *Pathways* for three academic years (nine classrooms). Seven teachers (three of them trainers) delivered *Pathways* twice (fourteen classrooms). Seventeen teachers (none trainers) delivered Pathways once (seventeen classrooms). All twelve of the teachers trained in Cycle 3 had only one opportunity to deliver *Pathways*. The others were four teachers in Cycle 1, and one who was trained in Cycle 2, but did not continue at the school.

4.3. Human subjects and IRB

We obtained Chicago Public Schools Research Review Board (CPS Project ID: 921) and USC IRB approval (USC: UP 1400287) before we collected data. We followed Federal rules which state that parental consent is not needed to evaluate programs that are part of the regular curriculum with existing data. We also followed CPS rules which state that parental consent is needed for new data collection.

[2] The U.S. Department of Education (2017) defines high-poverty schools as over 75% free/reduced-price lunch.

[3] The terminology is in flux. One possibility is to describe bi- or multi-racial or indigenous students and students of color as historically minoritorized rather than as minorities. The major point here is that most students did describe themselves as Latinx or African American/Black so that implies that though historically minoritized, these students are in the majority in the schools we worked with. The largest (New York), second (Chicago), and third (Los Angeles) largest school districts in the U.S. are also majority-minority, reporting about a 10% white enrollment. The U.S. Department of Education (2020) estimates that most African American and Latinx students but under 10% of white students are enrolled in schools with 75% or more minority enrollment and that about half of U.S. public schools have a 50% or higher minority enrollment. Hence, our schools represent a context common for students in urban schools and for students of color generally. At the same time, the Department of Education (2020) estimates that by fall 2029, 44% of public elementary and secondary students will be White, which implies that some school contexts are not like the schools in our sample (27% will be Latinx, 15% will be African American, 7% will be Asian/Pacific Islander, 6% will be of two or more racial-ethnic heritages).

[4] Due to researcher error, this information was not collected for Cycle 3 teachers.

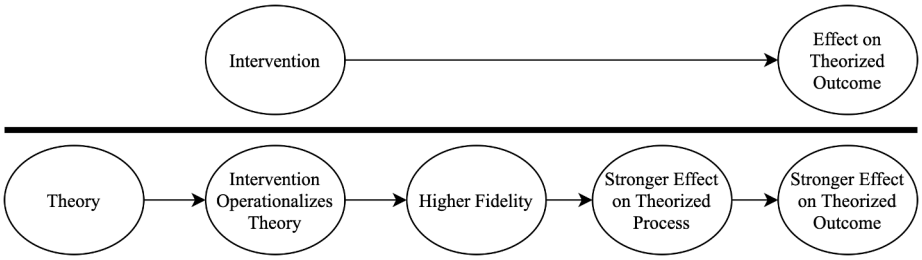

Fig. 2. Interventions Can Be Tests of Theories (Lower Panel), Though Often They Are Not (Upper Panel).

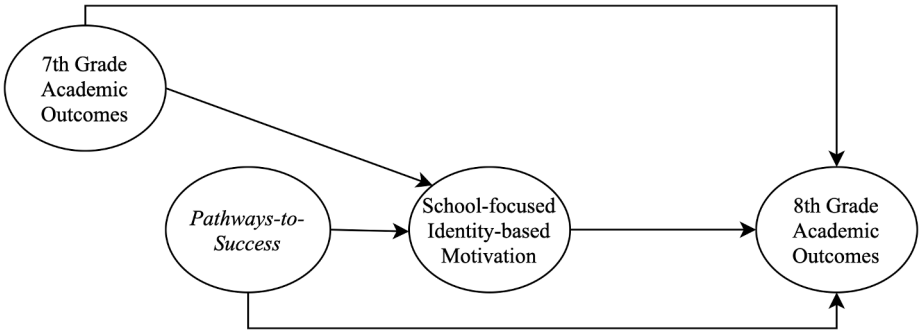

Fig. 3. Identity-based motivation is influenced by past experiences and participation in *Pathways*.

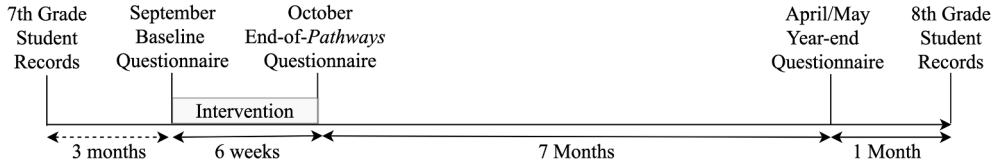

Fig. 4. Timing of Data Collection and *Pathways* Implementation.

4.4. Written consent

We explained our process to teachers. They provided written consent to be video-recorded implementing *Pathways* in their classrooms. We followed the CPS parental consent procedure (delivering letters and consent forms to each class for students to take home and bring back to school for pick-up). We analyzed the student surveys of the 88% of students who provided a signed form with the appropriate checkmark to affirm survey consent. Students who returned a (consenting or refusing) form received two movie tickets. Our consent rate parallels rates attained using other school-based high-effort strategies (mailing consent forms with report cards and calling parents, Esbensen et al., 1996; Pokorny, Jason, Schoeny, Townsend, & Curie, 2001).

4.5. Power analyses, stop-rules, transparency, and handling of missing data

We followed our planned stop-rules: stopping after engaging with 10 K-8 schools and the 8th-grade homeroom teachers in these schools. We analyzed all data. We provide the data we collected (CPS does not permit administrative data to be shared) online. Supplemental materials detail sample size per analysis. Data were missing if students were not present, had no parental consent, or skipped items. We took two steps to reduce the chance of losing participating children in our analyses. We impute child-level fidelity of receipt from classroom data. We use full information, multiple imputation techniques (described in our analysis plan)

5. Measures

5.1. Demographic information

We obtained gender, free/reduced price lunch status, and race-ethnicity data from administrative records as part of a data-sharing agreement between CPS and AIR to evaluate *Pathways*.

5.2. Academic outcomes

We obtained 7th- and 8th-grade grades, and course failure data from administrative records as part of a data-sharing agreement between CPS and AIR to evaluate *Pathways*. We computed a binary measure of course failure (0 = failed no classes, 1 = failed any class) and core course grade point average (GPA) with 4 = A, 3 = B, 2 = C, 1 = D, and 0 = F. Missing data were minimal, 3.6% of 7th-grade and 1.5% of 8th-grade GPA data were missing, as were 3.2% of 7th-grade and 0.4% of 8th-grade course failure data[5]. Missingness occurred when students were not attending CPS in 7th-grade or had left CPS before data were shared with us (information is not maintained when students leave).

[5] We used all information, even if incomplete – even with partial grade information, though we could not compute GPA, we could compute course failure if any of the available grades showed a failing grade.

Table 3
Iterative Development of *Pathways* Teacher-Training (TT) and Support.

Cycle Number	TT Length in Days	Who led?	When?		What?
1	2	PI	*Day 1*		*Pathways* as participants
			Day 2		Learn IBM theory and link to *Pathways*, practice implementing
			Weekly		Call in-support, one call-in per school
2	3	Teacher-educators with PI support	*Day 1*		Session 1–8 *Pathways* as participants
			Day 2	*AM*	Sessions 9–12 *Pathways* as participants
				PM	Learn IBM theory and link to *Pathways*
			Day 3		Practice implementing *Pathways* with structured feedback
			Weekly		Call-in support
3	3	Teacher-educators	*Day 1*		Session 1–8 *Pathways* as participants
			Day 2	*AM*	Sessions 9–12 *Pathways* as participants
				PM	Learn IBM theory and link to *Pathways*
			Day 3		Practice implementing *Pathways* with structured feedback
			Weekly		Call-in support, each trainer led a separate call

Note: TT = ETeacher training. We provide more detailed summaries of this information in supplemental materials.

Table 4
Teacher-driven changes.

Cycle Number	Changes to Implementation Manual	Changes to Materials	Changes to structure
1	Revised wordy or ambiguous content in implementation manual Added color coding and "call-out" boxes to the implementation manual	Simplified and laminated materials Replaced newsprint with PowerPoint and SMARTboard slides.	Codified the twice per week delivery format Divided session 4 into two sessions to facilitate student discovery. Added third training day
2	Further refinement of ambiguous and wordy content. Added more "call-out" boxes with teaching tips, modification ideas, examples of ways to structure processes of eliciting student knowledge, and sharing back	Launched website with implementation and video materials to support teachers	Revised session order to provide a more natural flow (former session 6 became session 8). Revised weekly call format, increasing structure. Developed a draft "educate-the-educator" manual for teacher educators
3	Further refinement of ambiguous or wordy content. Inclusion of student materials as examples.		Added email reminders to be received before weekly check-in calls. Refined the "educate-the-educator" manual to be more digestible to our teacher audience. Added PowerPoint, video, and other materials to support teacher-educators in covering the theoretical and empirical bases of *Pathways*.

5.3. Identity-based motivation

Next, we describe our measures of identity-based motivation. We share our full questionnaire in our Supplemental Materials.

Academic Possible Identity Scores. Students generated their possible identities and strategies in a sequence of prompts, as detailed in Fig. 5. We obtained a single score from all the text that they wrote using a machine-learning algorithm trained on a separate sample to score students' possible identities and strategies for their functional relevance to academics. The score is based on a support vector regression algorithm. The algorithm takes as its input a numeric representation of possible identities and strategies responses using Word2Vec. Word2Vec is a model that represents natural language use based on an existing corpus of news articles (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Fig. 6 summarizes the development of the algorithm in the separate development sample. As detailed in Supplemental Materials, to open the black box of machine coding, we looked descriptively at the spatial network of possible identity and strategy responses among students with higher scores and students with lower scores on the machine algorithm. Descriptively, students with higher scores tend to connect more facets of their possible identities and more of their strategies to school, elaborate more on them, and relate their interpersonal and off-track possible identities to school. O'Donnell and Oyserman detail these analyses elsewhere (O'Donnell & Oyserman, 2021).

Difficulty-as-Impossibility and Difficulty-as-Importance Scores. As detailed in Table 6, we used two 6-item, 5-point response (1 = *Strongly Disagree*, 5 = *Strongly Agree*) scales from Oyserman, Destin, and Novin (2015). An example item from the difficulty-as-impossibility scale ($\alpha = 0.86$) is: "When working on a task feels hard, that feeling means it's not for me." An example item from the difficulty-as-importance scale ($\alpha = 0.89$) is: "If a task is difficult, it is probably important for me to do well at it." As detailed in Supplemental Materials, we created a latent score for each measure.

Identity-based motivation latent score. As detailed in Supplemental Materials, we used the single machine-algorithm score along with the difficulty-as-importance and difficulty-as-impossibility latent scores to create an IBM latent score.

5.4. Self-regulatory competence

We collected exploratory data in AY15-16 and AY16-17 on whether engaging with *Pathways* increases student social-emotional competencies as operationalized by the Bailey and Halloran (2014) social-emotional competencies scale (15-item short-form; $\alpha = 0.92$, e.g., I understand why I do what I do. I pay attention when I do things. I am aware of how my mood affects the way I treat other people. I work hard; 1 = *Rarely*, 2 = *Occasionally*, 3 = *Frequently*, 4 = *Almost Always*). They validated the scale with over 20,000 students. We also used Aelenei, Lewis, and Oyserman (2017) 6-item 5-point response (1 = *Strongly Disagree* to *5* = *Strongly Agree*) goal investment scale (e.g., I put effort into attaining meaningful goals; $\alpha = 0.88$). As detailed in Supplemental Materials, we created latent constructs for each for analysis.

Table 5
Pathways by academic year, teacher, school, and training cycle.

Cycle	AY of Training	AY *Pathways* was delivered					
		2014–15		2015–16		2016–17	
		Schools	Teachers	Schools	Teachers	Schools	Teachers
1	2014–15	2	8	1	4	1	3
2	2015–16	—	—	3	7	2	6
3	2016–17	—	—	—	—	5	12
	Classrooms		8		11		21

Note: Numbers reflect the number of schools, teachers, and classrooms, respectively. In Cycle 1, one special education teacher taught a small standalone *Pathways* class. In other years, students receiving special education services were integrated with their peers and supported in their *Pathways* class.

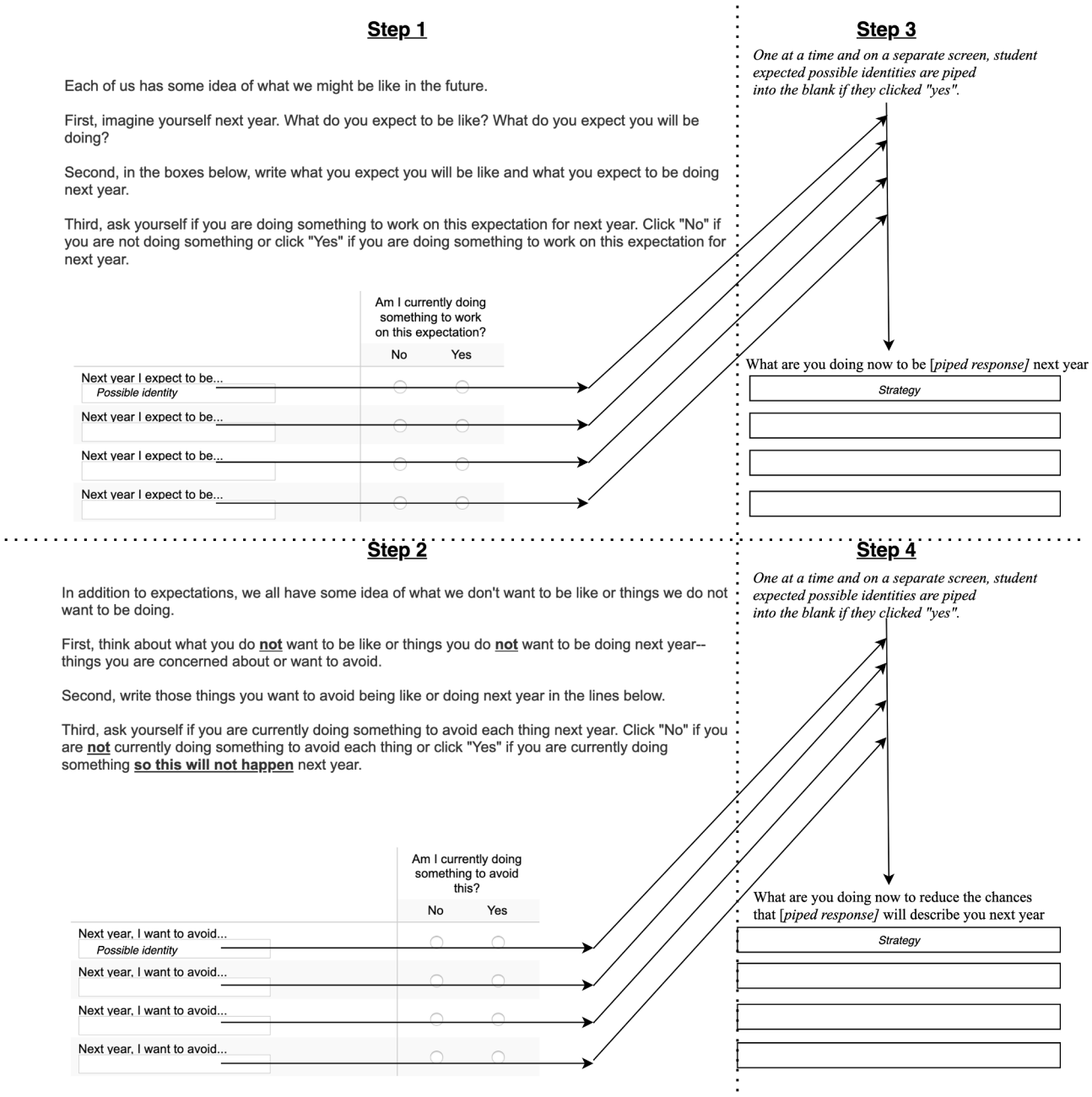

Fig. 5. Schematic representation of the possible identities and strategies measure, showing what students read in Qualtrics and the piping process which allowed them to see what they had previously written.

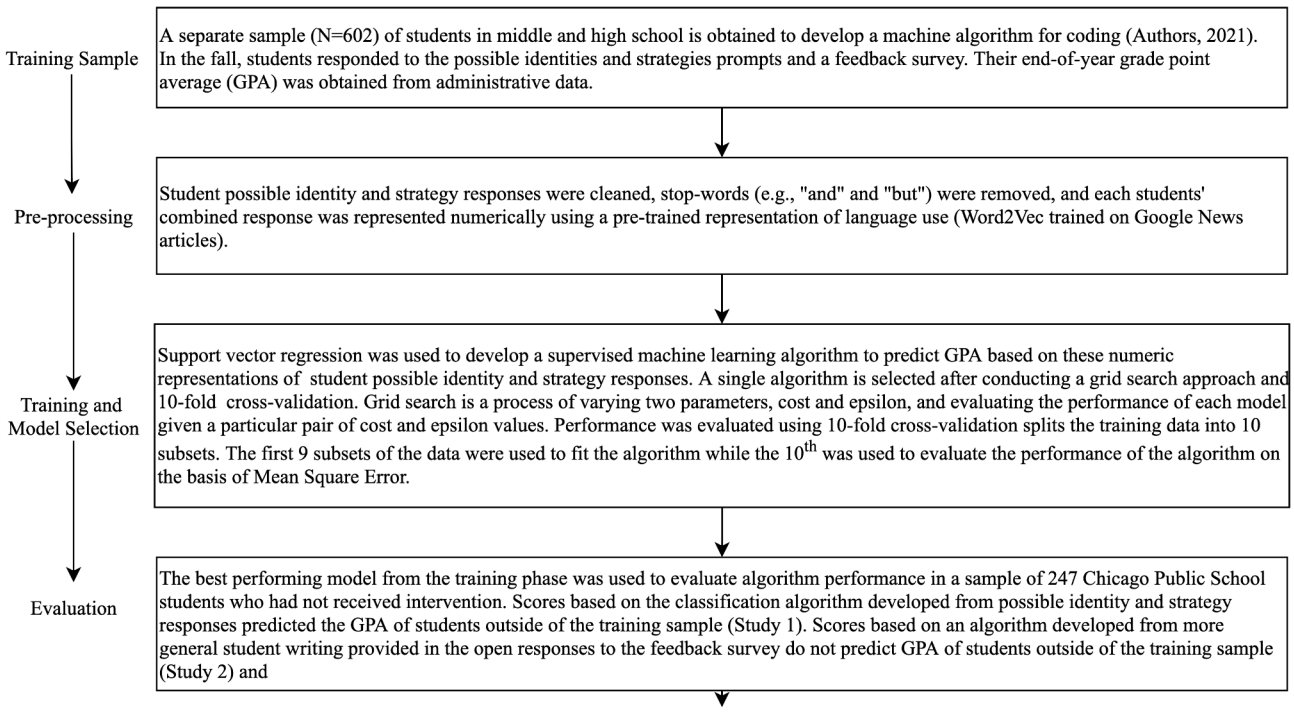

Fig. 6. Developing and validating a machine algorithm to score possible identities and strategies.

Table 6
Difficulty-as-importance and Difficulty-as-impossibility, Students Rated How Much They Agreed or Disagreed with 12 Statements Presented in Randomized Order.

Difficulty-as-importance statements	Difficulty-as-impossibility statements
If I'm working on a task that feels difficult, it means that the task is important.	*If I feel stuck on a task, it's a sign that my effort is better spent elsewhere.*
A sign that a task is important to me is how difficult it feels while working on it. If it feels difficult, it's important.	*If working on a task feels very difficult, that type of task may not be possible for me.*
Struggling to complete a task reminds me that the task is important.	*If a task feels too difficult, I should move on to something else.*
If a task is difficult, it is probably important for me to do well at it.	*When working on a task feels hard, that feeling means it's not for me.*
Tasks that feel difficult are important tasks.	*Finding a task really difficult tells me that I can't complete that task.*
If a task is difficult, it means that it's important for me.	*If a task feels really difficult, it may not be possible for me.*

Note: The instruction was: These next questions ask about your ideas about difficulty. Difficulty is often experienced by people working at, close to, or above their peak capacity. There is no right or wrong answer. Please indicate how much you agree or disagree with each of the following statements (1 = *Strongly Disagree* and 5 = *Strongly Agree*).

5.5. Fidelity

We coded each of the five aspects of fidelity (dosage α = 0.89, adherence α = 0.94, responsiveness α = 0.93, quality α = 0.75, and receipt α = 0.88). We coded dosage, adherence, and responsiveness and half of the components of quality from rater-coded video recordings of each session. We coded the other half of the components of quality and fidelity of receipt from the student-reported end-of-*Pathways* student feedback survey (collected within a week of completing *Pathways*). We averaged student-level information in obtaining a classroom-level fidelity score that we use to compare fidelity across cycles (H1). When we are focused on student-level effects (H2), we take another tack and keep student-level information at the student level, averaging in the aspects of fidelity that are based on classroom-level data. Since the two ways of thinking about fidelity are correlated, we only use one approach for each analysis. We summarize how we coded each aspect of fidelity in Table 7 and detail coding and averaging strategies in our Supplemental Materials).

5.6. Teacher quality inside and outside of Pathways

In the student feedback survey, students rated whether their teacher in each subject (math, science, English, history) was enthusiastic, warm, clear, and knowledgeable (1 = strongly disagree, 5 = strongly agree, α = 0.85). This allowed us to compare an aspect of teacher quality, teacher-driven climate, for teachers inside of *Pathways* and outside of *Pathways*. To preserve the independence of judgment in our analyses we compared how each teacher was rated within *Pathways* based on the ratings of their *Pathways* students describing them in *Pathways* to how each teacher was rated within their subject classes based on the ratings of the students in their subject class, excluding students who also had them for *Pathways*.

5.7. Analysis plan

Our planned analyses focus on the quality of the signal, both whether we can attain quality signal, and whether the quality of the signal matters for student IBM and hence their outcomes. We are not focused on comparing sub-groups of students in terms of their mean scores on our process measure of IBM or our outcome measures but provide these descriptive statistics in our Supplemental Materials.

H1: First we focused on fidelity in the classroom. We tested H1 descriptively in three ways. First, we compared the classroom-level data on fidelity available from the prior successful randomized trial (Oyserman et al., 2006) to parallel data in each of our 3 Cycles. Second, we compared classroom-level fidelity at each of our 3 cycles. Third, examined student experienced and classroom-level fidelity to see if below-threshold results were mostly in the first cycle.

H2: Then we looked at the relationship between the fidelity that students experience and their IBM and outcomes. We tested H2 using a structural equation model (SEM) with cluster robust standard

Table 7
Operationalization of five-factor fidelity scores.

Component of Fidelity Score	Source	
	Session-by-Session Video Observation	End of Intervention Student Feedback Survey
Adherence	Teacher activity checklist	–
Dosage	Task attempted checklist Delivery twice per week for 6 weeks	–
Responsiveness	Student behavior checklist CLASS-S Student Engagement Scales*	–
Quality	CLASS-S Teaching Quality Scales* The fullness of the take-home point Fluency	4-item teacher-driven climate scale 6-item teacher sensitivity scale 4-item peer-driven climate scale 6-item classroom support scale
Receipt	–	10-item Fidelity of Receipt scale

Note: *assessed twice each session.

errors in MPLUS (Muthén & Muthén, 1998, 2011). Our model estimates the indirect effects of the fidelity that students experience and 7th-grade academic outcomes on 8th-grade academic outcomes via their influence on our latent school-focused identity-based motivation variable. Our exploratory analyses of the effects of fidelity on social-emotional competence and goal investment used a similar approach and kept the measurement loadings for our IBM latent construct fixed to their values in our academic outcomes model. Our exploratory analyses of whether effects differ for students of color and white students follow the same approach as well.

5.8. Supplemental analyses

Subgroup Analyses. The nature of our study context and sample limits our ability to explore subgroup differences. As is typical of urban public schools, the students in our schools were predominately students of color (Black and especially Latinx). There were a relatively small number of white students (n = 136) and it is not feasible to identify other groups of students who might be similarly structurally advantaged using administrative demographic data. As an exploratory test of the robustness of our findings, we analyze our theoretical model twice, once with the full sample and a second time including only students of color (Black and Latinx). When the confidence intervals around parameter estimates in these two tests overlap, we conclude there is no meaningful difference in the estimates across the models with the full and restricted samples.

Comparison of teachers when they are and are not delivering Pathways. Qualitatively, in our teacher feedback interviews, we learned that some teachers wondered if they might be seen by their students as less expert in *Pathways* than they were in their regular classes and that many teachers believed that *Pathways* made them better teachers. We used two methods to address these questions. First, we report simple correlations of teacher-driven climate scores inside and outside of *Pathways* to assess the extent to which this aspect of teaching quality is similar across very different contexts. Second, we report average inside and outside of *Pathways* difference scores using a meta-analytic random-effects model (McShane & Böckenholt, 2017).

6. Results

6.1. H1: We can teach teachers to teach other teachers to implement Pathways with fidelity

We proceeded in three steps. First, we used classroom-level fidelity to compare our current results to the successful prior randomized test of *Pathways* (Oyserman et al., 2006). Only classroom-level dosage and adherence data could be retrieved from the randomized test, so we used classroom-level dosage and adherence data in our comparison. As depicted in Fig. 7, as we transitioned from the 2-day researcher-led training (Cycle 1) to the 3-day researcher-assisted teacher-led training (Cycle 2) and 3-day fully teacher-led training (Cycle 3) method, teachers attained dosage and adherence comparable to what was attained in the successful prior randomized test of *Pathways*.

Second, we used our 5-factor fidelity measure at the child level to compare our current results to our 60% rule-of-thumb fidelity threshold. In Fig. 8, we depict plots of the average fidelity students experienced at each cycle. We found that almost all (96.3%) students experienced fidelity above the 60% threshold, indeed, the median fidelity that students experienced was 75.2%. We repeated these analyses using classroom-level means and found comparable results, 97.5% of classrooms experienced *Pathways* at or above this 60% threshold, median classroom-level fidelity was 75.8%.

Third, we used our 5-factor measure of fidelity to unpack whether our development cycles sustained and supported fidelity. We display our results graphically in Fig. 9 as the cumulative percentage of students who experience each level of fidelity across cycles. As this graph shows, most students experienced fidelity in the 60% to 80% range. About one in four students (25.2%) experienced extremely high fidelity, above the 80% practical maximum. Even when students experienced lower fidelity, it tended to be close to the 60% threshold. Then we conducted a closer inspection of the 33 students with fidelity below the 60% threshold. This revealed that almost all (n = 30) were in classrooms whose teachers were trained in Cycle 1 (one was in a class whose teacher was trained in Cycle 2 and two were in a class whose teacher was trained in Cycle 3). The implication is that our development cycles supported fidelity. Supplemental Materials (Table S3) details fidelity by teacher-training cycle and academic year for classrooms (and within classrooms, minimum and maximum student-level fidelity).

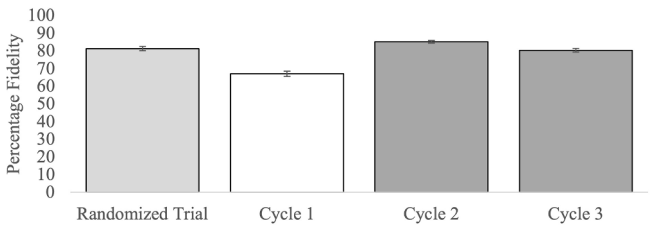

Fig. 7. Comparing Prior Implementation to Our Development Cycles: Mean Dosage and Adherence Fidelity Note. Whiskers are 95% confidence intervals. The left-most bar is the dosage and adherence fidelity attained with researcher-trained community members in a randomized trial with 5 days of training (Oyserman et al., 2006). The white bar is the dosage and adherence fidelity attained by teachers trained in Cycle 1 with 2 days of training (AY 2014–15), the two dark-grey bars represent teachers trained in Cycles 2 (AY 2015–16) and 3 (AY 2016–17) with 3 days of training.

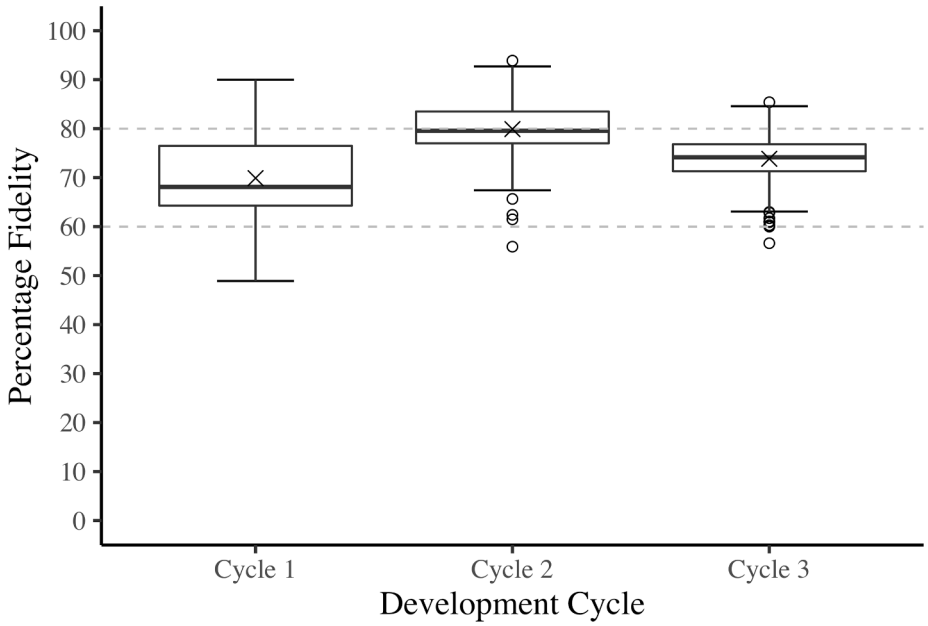

Fig. 8. Distribution of Child-level Fidelity by Cycle Note. Xs inside each box represent mean, lines inside each box represent median, box tops represent the top 75%, and box bottoms represent the bottom 25% of child-level fidelity. Dots are outliers, defined as child-level fidelity 1.5 times the difference between the highest and lowest quartiles.

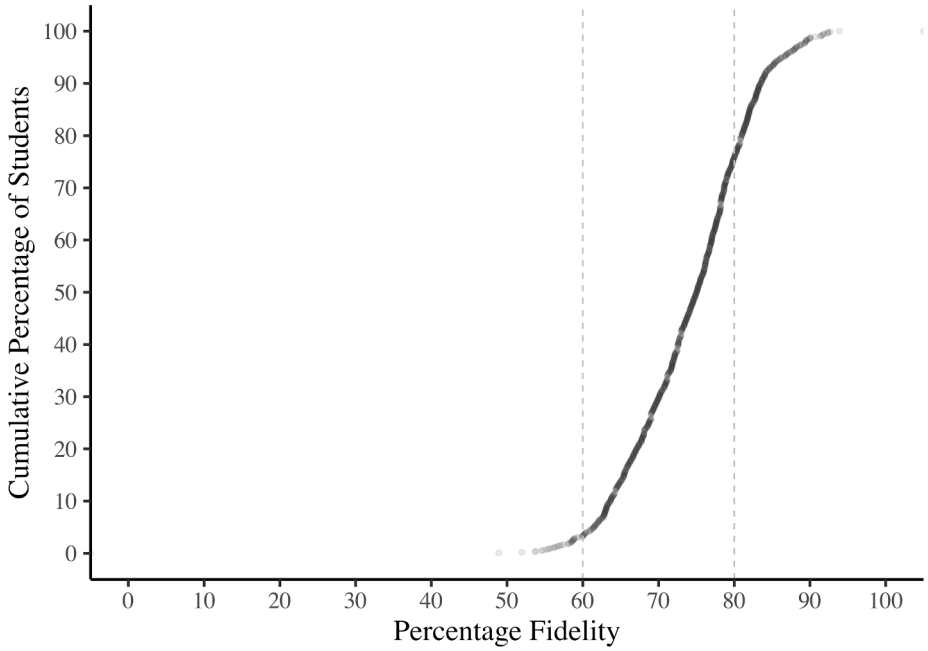

Fig. 9. Cumulative Percentage of Students as a Function of Fidelity, with 60% threshold and 80% Practical Maximum Marked.

6.2. H2: Higher fidelity pathways improvesacademic outcomes by improvingschool-focused identity-based motivation

We used student-experienced fidelity and our theoretical model (Fig. 3) as the basis of our analytic model (Fig. 10). As depicted in Fig. 10, by the end of the year, students scored higher in identity-based motivation ($p = .002$) if they experienced higher fidelity *Pathways*. Students with higher identity-based motivation had better academic outcomes at the end of 8th-grade ($p < .001$). Students experiencing *Pathways* with greater fidelity had better end-of-8th-grade academic outcomes because fidelity influenced their identity-based motivation ($ab = -0.109$, $p = .011$). Supporting our process-level prediction, the direct path from fidelity to academic outcomes was not significant ($p = .095$), implying that the process by which *Pathways* matters is captured by our latent identity-based motivation score rather than through other processes. Students' 7th-grade academic history affected their end-of-8th-grade academic outcomes ($p < .001$), and the better they were doing academically in 7th-grade, the higher their identity-based motivation ($p < .001$). It should come as no surprise that students with a history of success are more likely to succeed academically and those with a history of worse academic outcomes are less likely to do so. What is important to us is that experiencing *Pathways* with fidelity can reduce

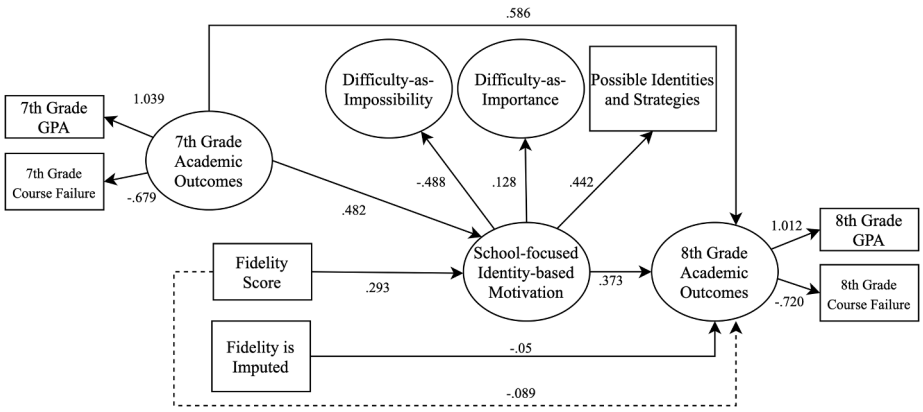

Fig. 10. Fidelity Influences Academic Outcomes via Identity-Based Motivation. *Note.* Dotted lines represent non-significant paths. Path coefficients are standardized. The model is a good fit to the data, $\chi^2(146) = 293.699$, $p < .001$. Fit indices reveal that our model recreated the sample correlation matrix over 95% of the time (RMSEA = 0.032, SRMR = 0.052). Both comparative fit indices indicate a good fit; the CFI suggests the model provides a 97.3% improvement and the TLI suggests it provides a 96.9% improvement over a baseline model in which all possible path coefficients are constrained to zero.

the influence of past academic failures on current outcomes. Students' academic past affects their future academic outcomes both directly ($p <$ .001) and indirectly by shaping their school-focused identity-based motivation ($ab = 0.179$, $p < .001$).

To get a sense of the size of this effect, we operationalized what moving from threshold (60%) to the practical maximum (80%) fidelity would yield in terms of academic outcomes. To do so, we multiplied the unstandardized indirect effect ($ab_{\text{unstandardized}} = 0.025$) by 20 (the difference between 60 and 80), obtaining a shift of half (0.50) of an *SD* in academic outcomes for a student in a classroom with the practical maximum fidelity rather than the threshold level.

We followed up with a reduced sample analysis to explore the possibility that effects are different for students of color (with only 135 white students, we do not have the statistical power to examine effects for white students). As detailed in Supplemental Materials, reduced sample results are within the confidence interval of full sample results. This indicates that we have no reason to assume that academic outcome results are different for students of color and white students.

6.3. H2 exploratory: Pathways bolsters self-regulatory competence by shaping identity-based motivation

In exploratory data collected only in the final two years (AY15-16, AY16-17), we explored the effect of fidelity on self-regulatory competence for students' social-emotional competence and goal investment scores. Model fit was adequate, our process models and fit indices are presented in Supplemental Materials. Because we only measured these variables in our second two years, our sample size and fidelity score variability are reduced. That said, higher fidelity influences identity-based motivation measures, and identity-based motivation measures influence self-regulatory competence. Children experiencing *Pathways* with higher fidelity scored higher in school-focused identity-based motivation ($p = .009$). By positively affecting identity-based motivation, experiencing *Pathways* with greater fidelity increases students' end of 8th-grade social-emotional competence ($ab = 0.137$, $p = .025$) and goal investment ($ab = 0.192$, $p = .013$). We then examined the direct path from higher fidelity to higher social-emotional competence ($p = .030$) and goal investment ($p = .593$). The direct path from fidelity to social-emotional competencies is significant. This implies that the effect of fidelity on social-emotional competence is not fully captured by our brief measures of identity-based motivation. For goal investment, the direct path is not significant, implying that the effects of fidelity are through our theorized IBM process. To understand our effects, we consider what moving from threshold (60%) to the practical maximum (80%) fidelity would yield in terms of increased self-regulatory competence. We found that a student in a classroom with the practical maximum fidelity had scores 0.88 and 0.68 of a standardized deviation higher for social-emotional competence and goal investment respectively than students at the threshold level. Given that these are latent constructs defined as having a mean of 0 and a standard deviation of 1, shifting from threshold to practical maximum has a substantial effect on student self-regulatory competencies.

We explored the possibility that these effects may be a function of student group membership by fitting the same models with data from our students of color only. In each model, though precision is lost, reduced sample results are within the confidence interval of full sample results, leaving us to infer that results are not contingent on group membership (full models, fit indices parameter estimates with confidence intervals are reported in Supplemental Materials)

6.4. Teachers provided a positive Teacher-Driven climate inside and outside of Pathways

Students rated their teachers as providing a positive teacher-driven climate, on a 1 (strongly disagree) to 5 (strongly agree) scale, on average agreeing that their teachers were enthusiastic, warm, clear, and knowledgeable in *Pathways* ($M = 4.18$, *95% CI*[4.09, 4.26]) and their subject classes ($M = 4.31$, *95 %CI*[4.23, 4.39]). Ratings were associated, $r(26) = 0.685$, $p < .001$. Teachers who were rated positively by their *Pathways* students also tended to be rated positively by their non-*Pathways* students. There was, however, a difference in inside and outside of *Pathways* ratings, though it was minuscule and heterogeneous (large random-effects heterogeneity statistic ($I^2 = 82.68$, 95% CI[78.10, 86.31]). This difference is depicted graphically in Fig. 11, when quantified, 82.6% of the variance in the inside-outside *Pathways* difference can be attributed to unmeasured differences between teachers rather than to *Pathways* itself. How teachers were experienced inside and outside of *Pathways* varied.

7. Discussion

We found support for our two predictions in our sample of over 1000 8th-graders in ten high-poverty majority-minority middle schools in which over three-quarters of students are eligible to free or reduced-price lunch given their family income and serving mostly students from historically minoritized groups. Our first prediction focused on addressing the training bottleneck. We predicted and showed that teachers, even when trained by other teachers, can implement *Pathways* with fidelity. Teachers trained by other teachers attained fidelity comparable to people trained in other ways. Our second prediction focused on assessing the theorized process. We predicted and showed that when students experience *Pathways* with fidelity, their school-focused identity-based motivation is bolstered. Higher school-focused identity-based motivation yields improved school success even when taking into account their prior academic trajectory. In this way, teachers can support their students to have more identity-based motivation. This support matters for two reasons. First, students with more identity-based motivation succeed. They are less likely to attain poor grades or fail courses.

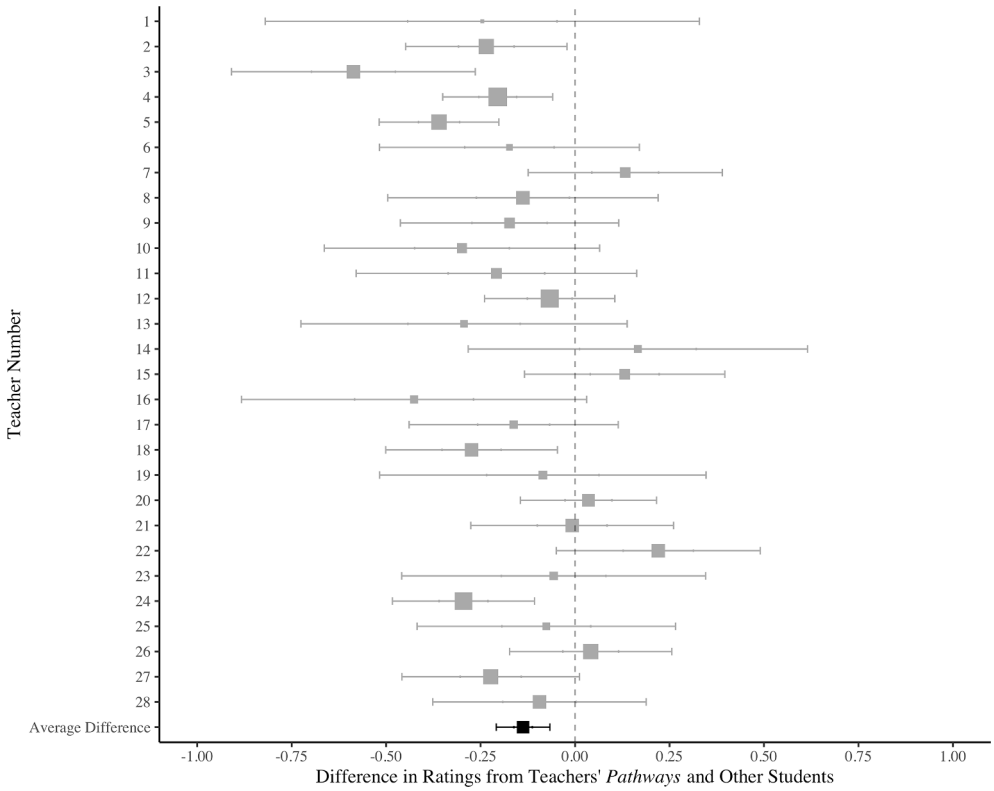

Fig. 11. The average difference in student ratings of teachers inside and outside of *Pathways* is qualified by substantial teacher-level heterogeneity in inside-outside difference scores.

Second, students with a history of course failures and bad grades otherwise have dampened identity-based motivation. Experiencing *Pathways* with fidelity mitigates this negative impact of the past on the future. We explored and found support for making a case that teachers who implement *Pathways* with fidelity also support their students in building social-emotional competence and goal investment. Our latent constructs of social-emotional competence and goal investment each focus on self-awareness, the *what* aspect of behavior. They do not focus on *how* to behave, choice of behavior is culturally constructed. We explored and found support for the possibility that by experiencing *Pathways* with fidelity, students were also supported in becoming better self-regulators by assessing effects on social-emotional competencies and willingness to invest in important goals.

Our results imply first, that *Pathways* is sustainable, and second, that identity-based motivation is a practical theory in the Lewin-ian (1951) sense. Teachers trained by other teachers can implement *Pathways* with fidelity in high-poverty schools serving mostly students from racially-ethnically minoritized backgrounds. Teachers who implement *Pathways* with fidelity help sustain supportive classroom environments and norms. Students in these classrooms have more identity-based motivation, and this bolsters their academic outcomes. Supporting our process model, the effects of fidelity on academic outcomes occur via their effect on identity-based motivation.

Our effects are small. But, for several reasons, they are meaningful. First, we are testing the effect of incremental improvement in *Pathways* implementation fidelity, not the effect of receiving *Pathways* versus school-as-usual. Second, like *Pathways* itself, our training is short and scalable, and sustainable (teachers can train other teachers). Third, we document that intervention effects occur by affecting identity-based motivation, supporting our process model. Third, as we detail next, our results connect with and advance the literature on self and motivation, highlight an implied but not tested connection between identity and social norms, and highlight how interventions can test the theories they operationalize.

7.1. Self, motivation, social norms, and academic outcomes

A central puzzle in the motivation literature is the aspiration-attainment gap in which students want to do well but do not get the needed support and so start too late, persist too little, get distracted, and even disengage. Identity-based motivation theory predicts that school and classroom contexts shape the size of this gap in several ways. School and classroom contexts can signal that the future is far and disconnected from the present (Nurra & Oyserman, 2018), that school is irrelevant to future goals (Harackiewicz, Rozek, Hulleman, & Hyde, 2012; Hulleman & Cordray, 2009). School and classroom contexts can also signal to students that their school goals are unattainable (Hulleman, Godes, Hendricks, & Harackiewicz, 2010), or that they lack the needed ability (intellectual or academic, Sisk et al., 2018). School and classroom contexts can signal that people like themselves do not belong and are unlikely to succeed in school (Cohen, Purdie-Vaughns, & Garcia, 2012; Schunk & Ertmer, 2000).

Identity-based motivation theory predicts that each of these context effects is due in part to how they shape student perceptions of what is normative (typical or desirable) for students like themselves. Identity-based motivation theory predicts that students are sensitive to normative information because they prefer to act and make sense of their experiences in identity-congruent ways. Using this identity-based motivation lens, each of these context-effects is about the future-focused and social aspects of identity. Thus, students may implicitly infer from classroom norms that it is typical or desirable for students like themselves to feel each of the following: My future is far, not relevant to now. School is irrelevant to the person I want to become. I do not expect to attain success at school. I lack the ability needed to succeed in school. School is not a place where I belong.

Given this common core, interventions targeting each of these aspects (connection, relevance, expectations, effort, and belonging) are compatible with the identity-based motivation prediction that these aspects of identity are context-sensitive, malleable, and matter for

student school-focused motivation and persistent engagement. Moreover, our results support the social grounding of identity-based motivation theory. By showing that the fidelity with which teachers implement classroom intervention matters, we concretize what has been implied by past work on the effect of social norms on individual beliefs (Cialdini & Goldstein, 2004; Paluck & Shepherd, 2012; Paluck, 2009; Sherif, 1936).

7.2. Interventions as tests of theories

Positive intervention results require evidence of the process for results to be a theory test. Critical reviews articulate three reasons that interventions rarely yield such tests (Durlak et al., 2010, 2011; Leeuw & Vaessen, 2009; Prestwich et al., 2015). First, intervention reports often fail to document fidelity. Second, interventions may not even claim a theoretical base (or report building on several theories without clarifying how they fit together). Third, those interventions that do claim a theoretical base often cannot test their claims for several reasons: they do not fully specify how manualized activities and techniques operationalize the theory and they do not provide measures of the theorized process constructs these manualized activities are to influence. Together, these omissions yield ambiguity as to whether an intervention tested the theory on which it was ostensibly based and undermine confidence as to whether results are replicable.

Our development process provides a roadmap for using interventions as theory tests. The *Pathways* intervention is manualized. Session activities operationalize each component of identity-based motivation theory. Across sessions, students experience multiple operationalizations of each identity-based motivation component. We obtained brief measures of these core components and created a latent measure so we could test that the theorized process. We documented that students on average experience fidelity above a plausible threshold of 60%, that students' academic outcomes are better when their identity-based motivation is higher, and that higher identity-based motivation is a function of experienced fidelity. We make two inferences from these results. First, students are doing better academically in part because their contexts have changed. Second, identity-based motivation is a practical theory in the Lewin (1951) sense of predicting a process with real-world consequences.

7.3. Practical implications

We focus on two practical implications from our methods and results. First, regarding our methods, we provide a template for future research. We worked with teachers to intervene at the whole-classroom level and assessed fidelity and attained brief measures of our theorized process. Our method of scoring fidelity uses an operationalization of the five fidelity components (dosage, adherence, quality, responsiveness, and receipt) assessed with a structured set of content-codes and student reports. We were able to compare fidelity across people and methods of training and use fidelity scores and brief measures of our theorized process to test whether the clearer the signal, the more positive the result.

Scoring fidelity and measurement of the theorized process can clarify sources of heterogeneity. If scored fidelity is below an agreed-upon threshold, heterogeneity likely entails gaps between planned and actual intervention. Researchers can look at when this occurs and what obstacles might be, potentially increasing fidelity across iterative cycles as we did between Cycles 1 and 2. Researchers could diagnose heterogeneity in results as due to variability in implementation if they find that implementation fidelity varies from under-threshold to over-threshold across studies. Heterogeneity in implementation fidelity would imply that variability might be due to training or to context-based differences in the viability of the intervention. Researchers should focus on hidden moderators or other limitations of the theory as the source of heterogeneity in effects if fidelity is at the agreed-upon threshold but does not always affect the process measures. For example, it might be the case that theory-predicted processes only work for some children or in kinds of schools. Examination of heterogeneity in results when the intervention implementation attains fidelity would allow for revising the theory (e.g., revealing that the theory describes processes only relevant to low-achieving students or only to students in low-achieving schools, Allcott, 2015; Borman et al., 2018; Sisk et al., 2018).

Second, regarding results, we interpret our positive findings to imply that the more *Pathways* is implemented and experienced with fidelity, the more students repeatedly model productive school-focused identity-based motivation for one another. Teachers do their part by attending to these behaviors, increasing the chances that students infer that these are socially normative (Ma et al., 2018; Sai, Liu, Li, Compton, & Heyman, 2020; Zhao et al., 2019). A practical implication of our results is that intervening at the classroom level is a viable way to change norms. Though we focused on teachers as facilitators, alternative routes to invoking norm change could work as well. For example, it might be possible to implement *Pathways* with peer influencers by giving them tools to relay *Pathways* messages to their peers. This approach requires identifying students who are salient and well connected and intervening with them to teach them to spread core messages (e.g., Paluck, Shepherd, & Aronow, 2016; Stephens, Townsend, Hamedani, Destin, & Manzo, 2015; Valente, 2012). Like other peer-driven interventions, *Pathways* leverages peer interaction. In the case of classroom-based *Pathways*, students engage in classroom-wide shared activities, and teachers publicly reinforce their engagement and active participation, as well as their personalization of core *Pathways* messages. Other peer-driven interventions require either census of the school to identify influencers (Paluck & Shepherd, 2012) or researcher skill in choosing attractive student models to provide compelling messages to other students (Stephens, Hamedani, & Destin, 2014). Our classroom-level *Pathways* intervention does not require either of these.

8. Limitations

Like all study designs, our design has both strengths and limits. In terms of strengths, we built on prior randomized trial results with a community member-implemented version of *Pathways* (Oyserman et al., 2006) and a manual that detailed how community members working in pairs could implement *Pathways* with fidelity (Oyserman, 2015). Our three-cycle iterative development design allowed us to document that we could shift to teacher implementation and teacher-led training and show, via scored implementation fidelity, that this change preserved fidelity at or above our agreed-upon fidelity threshold. Our design also allowed us to document that incrementally increasing implementation fidelity above this threshold and toward a likely practical maximum of 80% is worthwhile– it improves academic outcomes by enhancing student identity-based motivation. At the same time, our design limitation is that effect sizes must be interpreted differently from randomized trials. Effect sizes reflect the effect of incrementally higher fidelity beyond threshold rather than of being randomly assigned to *Pathways* or school-as-usual. Our development design means that it is not possible to directly compare our effect sizes with the effects of randomized trials because we are looking only at the effect of an incremental increase in fidelity amongst treated students.

That said, our effect is meaningful relative to other intervention attempts. Of the 43 proposals funded by the U.S. Department of Education's Investing in Innovation (i3) grants through 2018 focused on middle school intervention, only 47% reported a positive effect and the mean effect of the 67% that reported any effect at all was 0.062 (Boulay et al., 2018). This effect is within the bounds of the median average effect for all universal prevention programs for school-aged children (various outcomes, Tanner-Smith, Durlak, & Marx, 2018). These studies often examined whole-school and whole-year interventions. Our observed effect of moving from a 60% threshold to likely maximum 80% fidelity is impressive because it is in the upper half of what can be expected when comparing the effect of receiving versus not receiving intervention at all.

Moreover, our design allowed us to address a key feature of scalability, which is whether teachers can provide the nexus rather than being dependent on a researcher-led training process. We focused on a large urban school district and our results may comfortably apply to these contexts. A limitation of our design is that we cannot know whether changing to a dramatically different, rural, context matters. We are currently in the process of testing our model in rural settings and, in a few years, we hope to be able to begin to document what our current data cannot, which is scalability in rural settings.

9. Conclusions

Most children attend schools in urban areas, and most school children are from diverse racial-ethnic backgrounds (the majority are minorities). Indeed, 80% of Latinx and African American children attend schools in which over 50% of the student body is a member of a minoritized group (de Brey, Musu, McFarland, Wilkinson-Flicker, Diliberti, Zhang, & Wang, 2018). Our Chicago schools and students are like the schools and students in other large urban areas, and so our results may generalize. Our results suggest that experiencing the *Pathways* intervention with incrementally higher fidelity enhances school-focused identity-based motivation, bolstering their academic outcomes. Schools and teachers can shape education contexts by implementing *Pathways*.

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.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cedpsych.2021.101993.

References

ACT (2018). The condition of college and career readiness: National 2018. MR00912. Retrieved 09-24-2019 https://www.act.org/content/dam/act/unsecured/documents/cccr2018/National-CCCR-2018.pdf.

Aelenei, C., Lewis, N. A., Jr., & Oyserman, D. (2017). No pain no gain? Social demographic correlates and identity consequences of interpreting experienced difficulty as importance. *Contemporary Educational Psychology, 48*, 43–55.

Allcott, H. (2015). Site selection bias in program evaluation. *Quarterly Journal Economics, 130*(3), 1117–1165.

Bailey, P., & Halloran, C. (2014). *Technical report on the psychometric properties of the student self-report of social and emotional competence, grades 7 and 10*. Washington, D. C.: American Institute of Research.

Beal, S. J., & Crockett, L. J. (2010). Adolescents' occupational and educational aspirations and expectations: Links to high school activities and adult educational attainment. *Developmental Psychology, 46*(1), 258–265.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., … Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology, 23*(5), 443–451.

Borman, G., Grigg, J., Rozek, C., Hanselman, P., & Dewey, N. (2018). Self-affirmation effects are produced by school context, student engagement with the intervention, and time: Lessons from a district-wide implementation. *Psychological Science, 29*(11), 1773–1784.

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., & Rimdzius, T. (2018). *The Investing in Innovation Fund: Summary of 67 evaluations*. U.S. Department of Education.

Brunello, G., Fort, M., Schneeweis, N., & Winter-Ebmer, R. (2016). The causal effect of education on health: What is the role of health behaviors? *Health Economics, 25*(3), 314–336.

Cialdini, R., & Goldstein, N. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology, 55*, 591–621.

Clarke, I. E., & MacCann, C. (2016). Internal and external aspects of self-handicapping reflect the distinction between motivations and behaviours: Evidence from the Self-handicapping Scale. *Personality and Individual Differences, 100*, 6–11.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cohen, G. L., Purdie-Vaughns, V., & Garcia, J. (2012). An identity threat perspective on intervention. In M. Inzlicht, & T. Schmader (Eds.), *Stereotype threat: Theory, process, and application* (pp. 280–296). Oxford University Press.

Conti, G., Heckman, J., & Urzua, S. (2010). The education-health gradient. *American Economic Review, 100*(2), 234–238.

Cowan, B. (2018). Sources of bias in teenagers' college expectations. *Social Science Quarterly, 99*(1), 136–153.

Crosse, S., Williams, B., Hagen, C. A., Harmon, M., Ristow, L., DiGaetano, R.,... & Derzon, J. (2011). Prevalence and implementation fidelity of research-based prevention programs in public schools. Final Report. Washington, D.C.: Office of Planning, Evaluation and Policy Development, U.S. Dep't of Education. http://files.eric.ed.gov/fulltext/ED529062.pdf.

Dane, A., & Schneider, B. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*(1), 23–45.

de Brey, C., Musu, L., McFarland, J., Wilkinson-Flicker, S., Diliberti, M., Zhang, A.,... & Wang, X. (2019). Status and trends in the education of racial and ethnic groups 2018. (NCES 2019-038). U.S. Dep't of Education. National Center for Education Statistics. https://nces.ed.gov/programs/raceindicators/.

Destin, M., & Oyserman, D. (2009). From assets to school outcomes: How finances shape children's perceived possibilities and intentions. *Psychological Science, 20*(4), 414–418.

Duncan, G. J., & Murnane, R. J. (2011). Introduction, the American Dream then and now. In G. J. Duncan, & R. J. Murnane (Eds.), *Whither opportunity?: Rising inequality, schools, and children's life chances* (pp. 3–26). Russell Sage Foundation.

Durlak, J., & DuPre, E. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327–350.

Durlak, J., Weissberg, R., Dymnicki, A., Taylor, R., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405–432.

Durlak, J., Weissberg, R., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology, 45*, 294–309.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. (2003). A review of research on fidelity of implementation. *Health Education Research, 18*, 237–256.

Eddy, J. M., Martinez, C. R., Grossman, J. B., Cearley, J. J., Herrera, D., Wheeler, A. C., … Harachi, T. W. (2017). A randomized controlled trial of a long-term professional mentoring program for children at risk: Outcomes across the first 5 years. *Prevention Science, 18*(8), 899–910.

Elmore, K., & Oyserman, D. (2012). If 'we' can succeed, 'I' can too: Identity-based motivation and gender in the classroom. *Contemporary Educational Psychology, 37*, 176–185.

Esbensen, F., Piper, D. E., Vogel, R., West, J., Arboit, K., & Harris, L. (1996). Active parental consent in school-based research: An examination of ethical and methodological issues. *Evaluation Review, 20*, 737–753.

Fisher, O., & Oyserman, D. (2017). Assessing interpretations of experienced ease and difficulty as motivational constructs. *Motivation Science, 3*(2), 133–163. https://doi.org/10.1037/mot0000055.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*(102), 74–78.

Harackiewicz, J., Rozek, C., Hulleman, C., & Hyde, J. (2012). Helping parents to motivate adolescents in mathematics and science: An experimental test of a utility-value intervention. *Psychological Science, 23*, 899–906.

Herschell, A. D., Kolko, D. J., Baumann, B. L., & Davis, A. C. (2010). The role of therapist training in the implementation of psychosocial treatments: A review and critique with recommendations. *Clinical Psychology Review, 30*(4), 448–466.

Hirt, E., & McCrea, S. (2009). Man smart, woman smarter? Getting to the root of gender difference in self-handicapping. *Social and Personality Psychology Compass, 3*, 260–274.

Horowitz, E., Sorensen, N., Yoder, N., & Oyserman, D. (2018). Teachers can do it: Scalable identity-based motivation intervention in the classroom. *Contemporary Educational Psychology, 54*, 12–28.

Horowitz, E., Oyserman, D., Dehghani, M., & Sorensen, N. (2020). Changing academic trajectories: Which changes in possible selves matter? *Journal of Adolescence, 79*, 26–38.

Hulleman, C., & Cordray, D. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*, 88–110.

Hulleman, C., Godes, O., Hendricks, B., & Harackiewicz, J. (2010). Enhancing interest and performance with utility value intervention. *Journal of Educational Psychology, 102*, 880–895.

Jacob, B., & Linkow, T. (2011). Educational expectations and attainment. In G. Duncan, & R. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 133–162). Sage.

Jones, E. E., & Berglas, S. (1978). Control of attributions about the self through self-handicapping strategies: The appeal of alcohol and the role of underachievement. *Personality and Social Psychology Bulletin, 4*(2), 200–206.

King-Sears, M. E., Walker, J. D., & Barry, C. (2018). Measuring teachers' intervention fidelity. *Intervention in School and Clinic, 54*(2), 89–96.

Leeuw, F., & Vaessen, J. (2009). Carefully articulate the theories linking interventions to outcomes. Impact evaluations and development: Network of Networks for Impact Evaluation (NONIE) guidance on impact evaluation (Working Paper 57490). http://documents.worldbank.org/curated/en/411821468313779505/pdf/574900WP0Box351BLIC10nonie1guidance.pdf.

Lewin, K. (1951). Problems of research in social psychology. In D. Cartwright (Ed.), *Field theory in social science: Selected theoretical papers* (pp. 155–169). Harper & Row.

Lewis, N. A., Jr, & Yates, J. F. (2019). Preparing disadvantaged students for success in college: Lessons learned from the preparation initiative. *Perspectives on Psychological Science, 14*(1), 54–59.
Ma, F., Heyman, G., Jing, C., Fu, Y., Compton, B., Xu, F., & Lee, K. (2018). Promoting honesty in young children through observational learning. *Journal of Experimental Child Psychology, 167*, 234–245.
McFarland, J., Hussar, B., Wang, X., Zhang, J. Wang, K., Rathbun, A, … & Ossolinski, M. (2018). The Condition of Education 2018. (NCES 2018-144). U.S. Department of Education. National Center for Education Statistics. https://nces.ed.gov/pubs2018/2018144.pdf.
McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: A review of current efforts. *American Psychologist, 65*(2), 73–84.
McShane, B. B., & Böckenholt, U. (2017). Single paper meta-analysis: Benefits for study summary, theory-testing, and replicability. *Journal of Consumer Research, 43*(6), 1048–1063.
Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (pp. 3111–3119). Redhook, NY: Curran Associates Inc.
Mowbray, C., Holter, M. C., Teague, G., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation, 24*, 315–340.
Muthén, L. K., & Muthén, B. O. (1998-2011). Mplus user's guide. 8th ed. Muthén & Muthén.
Nurra, C., & Oyserman, D. (2018). From future self to current action: An identity-based motivation perspective. *Self and Identity, 17*(3), 343–364.
O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*, 33–84.
O'Donnell, S. C., & Oyserman, D. (2021). Bridging possible identities support academic attainment. Unpublished Manuscript. In preparation.
O'Donnell, S. C., Yan, V. X., Bi, C., & Oyserman, D. (2021). Is difficulty mostly about impossibility? What difficulty implies may be culturally variant. *Under editorial Review*. Submitted for publication.
Oyserman, D. (2009). Identity-based motivation: Implications for action-readiness, procedural-readiness, and consumer behavior. *Journal of Consumer Psychology, 19*(3), 250–260.
Oyserman, D. (2015). *Pathways to success through identity-based motivation*. Oxford University Press.
Oyserman, D., Bybee, D., & Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology, 91*, 188–204.
Oyserman, D., Elmore, K., & Smith, G. (2012). Self, self-concept and identity. In M. Leary, & J. Tangney (Eds.), *Handbook of Self and Identity* (pp. 69–104). Guilford Press.
Oyserman, D., Destin, M., & Novin, S. (2015). The context-sensitive future self: Possible selves motivate in context, not otherwise. *Self and Identity, 14*(2), 173–188.
Oyserman, D., Lewis, N., Jr., Yan, V., Fisher, O., O'Donnell, S. C., & Horowitz, E. (2017). An identity-based motivation framework for self-regulation. *Psychological Inquiry, 28*, 139–147.
Oyserman, D., Novin, S., Elmore, K., Fisher, O., & Smith, G. C. (2018). Guiding people to interpret their experienced difficulty as importance highlights their academic possibilities and improves their academic performance. *Frontiers in Psychology, 9*. article 781.
Oyserman, D. (2007). Social identity and self-regulation. In A. Kruglanski, & T. Higgins (Eds.), *Handbook of Social Psychology* (2nd ed., pp. 432–453). Guilford.
Oyserman, D. (2013). Not just any path: Implications of identity-based motivation for disparities in school outcomes. *Economics of Education Review, 33*, 179–190.
Oyserman, D., Bybee, D., Terry, K., & Hart-Johnson, T. (2004). Possible selves as roadmaps. *Journal of Research in Personality, 38*(2), 130–149.
Oyserman, D., Gant, L., & Ager, J. (1995). A socially contextualized model of African American identity: possible selves and school persistence. *Journal of Personality and Social Psychology, 69*, 1216–1232.
Oyserman, D., Terry, K., & Bybee, D. (2002). A possible selves intervention to enhance school involvement. *Journal of Adolescence, 25*(3), 313–326.
Paluck, E. L. (2009). What's in a norm? Sources and processes of norm change. *Journal of Personality and Social Psychology, 96*, 594–600.
Paluck, E., & Shepherd, H. (2012). The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network. *Journal of Personality and Social Psychology, 103*(6), 899–915.
Paluck, E., Shepherd, H., & Aronow, P. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences, 113*, 566–571.
Pokorny, S., Jason, L., Schoeny, M., Townsend, S., & Curie, C. (2001). Do participation rates change when active consent procedures replace passive consent. *Evaluation Review, 25*, 567–580.
Prestwich, A., Webb, T., & Conner, M. (2015). Using theory to develop and test interventions to promote changes in health behaviour: Evidence, issues, and recommendations. *Current Opinion in Psychology, 5*, 1–5.
Resnick, B., Bellg, A., Borrelli, B., De Francesco, C., Breger, R., … Ogedegbe, G. (2005). Examples of implementation and evaluation of treatment fidelity in the BCC studies: Where we are and where we need to go. *Annals of Behavioral Medicine, 29*(2), 46–54.
Rowan, B. (2011). Intervening to improve the educational outcomes of students: Lessons from recent work in high poverty schools. In G. Duncan, & R. Murnane (Eds.), *Whither opportunity?: Rising inequality, schools, and children's life chances* (pp. 523–538). Sage.
Sai, L., Liu, X., Li, H., Compton, B. J., & Heyman, G. D. (2020). Promoting honesty through overheard conversations. *Developmental Psychology, 56*(6), 1073–1079.
Schunk, D., & Ertmer, P. (2000). Self-regulation and academic learning: Self-efficacy enhancing interventions. In *Handbook of self-regulation* (pp. 631–649). Academic Press.
Sherif, M. (1936). *The psychology of social norms*. Ny: Harper.
Sisk, V., Burgoyne, A., Sun, J., Butler, J., & Macnamara, B. (2018). To what extent and under which circumstances are growth mind-sets important to academic achievement? *Two meta-analyses. Psychological Science, 29*(4), 549–571.
Smith, G. C. & Oyserman, D. (2015). Just not worth my time? Experienced difficulty and time investment. Social Cognition, 33, 85–103. https://doi.org/10.1521/soco.2015.33.2.1.
Smith, G. C., James, L., Varnum, M., & Oyserman, D. (2014). Give up or get going? Productive uncertainty in uncertain times. *Self and Identity, 13*(6), 681–700.
Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science, 10*(2), 176–199.
Stephens, N., Hamedani, M., & Destin, M. (2014). Closing the social class achievement gap: A diversity education intervention improves first-generation students' academic performance and all students' college transition. *Psychological Science, 25*, 943–953.
Stephens, N., Townsend, S., Hamedani, M., Destin, M., & Manzo, V. (2015). A difference-education intervention equips first-generation college students to thrive in the face of stressful college situations. *Psychological Science, 26*(10), 1556–1566.
Tanner-Smith, E. E., Durlak, J. A., & Marx, R. A. (2018). Empirically based mean effect size distributions for universal prevention programs targeting school-aged youth: A review of meta-analyses. *Prevention Science, 19*(8), 1091–1101.
Valente, T. W. (2012). Network interventions. *Science, 337*(6090), 49–53.
Vivalt, E. (2020). How much can we generalize from impact evaluations? *Journal of the European Economics Association, 18*(6), 3045–3089.
Weingarten, C., Rabago, J., Reynolds, J., Gates, K., Yanagida, E., & Baker, C. (2018). Examining the utility of a train-the-trainer model for dissemination of sexual violence prevention in schools. *Child Abuse & Neglect, 80*, 70–79.
Wolniak, G., Davis, L., Williams, T., & Casano, K. (2016). The State of Our Nation's Youth: 2016. Horatio Alger Association of Distinguished Americans, Inc. downloaded 09/24/19 https://www.horatioalger.org/wp-content/uploads/2016/09/HA-2016-SONY-FINAL.pdf.
Xu, J. (2006). Gender and homework management reported by high school students. *Educational Psychology, 26*, 73–91.
Xu, J. (2007). Middle school homework management. *Educational Psychology, 27*(2), 173–189.
Zhao, L., Chen, L., Sun, W., Compton, B., Lee, K., & Heyman, G. (2019). Young children are more likely to cheat after overhearing that a classmate is smart. *Developmental Science*, Article e12930.