Current Psychology (2024) 43:7547–7563
https://doi.org/10.1007/s12144-023-04871-z

Networks and directed acyclic graphs: Initial steps to efficiently examine causal relations between expectancies, values, and prior achievement

Hye Rin Lee[1,2] · Xin Tang[3] · Daniela Alvarez-Vargas[2] · Ji Seung Yang[4] · Drew Bailey[2] · Sandra Simpkins[2] · Nayssan Safavian[2] · Hanna Gaspard[5] · Katariina Salmela-Aro[6] · Julia Moeller[7] · Jacquelynne S. Eccles[2] · Allan Wigfield[4]

Accepted: 9 June 2023 / Published online: 13 July 2023

Abstract

The present study explored a set of plausible directed acyclic graphs (DAGs) of constructs involved in Situated Expectancy–Value Theory (SEVT) using cross-sectional data. To do so, three datasets ($n=1,540$; 1,867; and 103) with expectancy, values, and prior achievement constructs were used. First, networks showed a consistent magnitude of associations between SEVT constructs across datasets. Expectancy was associated with achievement ($r=.21$; 0.29; 0.37) and intrinsic value ($r=.35$; 0.68; 0.60); intrinsic value was associated with attainment value ($r=.31$; 0.79; 0.61). We reason through a set of plausible DAGs, or hypothesized directional models. Across datasets, DAGs revealed that prior achievement relates to expectancy, intrinsic value relates to expectancy, and utility relates to attainment value. This study highlights that DAGs can be used as a step to help triangulate, in conjunction with experimental and longitudinal work, on the underlying processes shaping relations between constructs.

Keywords Directed acyclic graphs · Mathematics · Motivation · Networks · Situated Expectancy–Value Theory · Values

Hye Rin Lee and Xin Tang contributed equally to this work.

✉ Hye Rin Lee
hyerin@udel.edu

Xin Tang
xin.tang@sjtu.edu.cn

Daniela Alvarez-Vargas
dalvare5@uci.edu

Ji Seung Yang
jsyang@umd.edu

Drew Bailey
dhbailey@uci.edu

Sandra Simpkins
simpkins@uci.edu

Nayssan Safavian
nayssan.safavian@uci.edu

Hanna Gaspard
hanna.gaspard@tu-dortmund.de

Katariina Salmela-Aro
katariina.salmela-aro@helsinki.fi

Julia Moeller
julia.moeller@uni-leipzig.de

Jacquelynne S. Eccles
jseccles@uci.edu

Allan Wigfield
awigfiel@umd.edu

1 College of Education & Human Development, University of Delaware, 106 Alison Hall West, Newark, DE 19716, USA

2 School of Education, University of California, Irvine, CA, USA

3 School of Education, Shanghai Jiao Tong University, Shanghai, China

4 Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD, USA

5 Center for Research on Education and School Development, TU Dortmund University, Dortmund, Germany

6 Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

7 Department of Educational Psychology, Leipzig University, Leipzig, Germany

Although many theorists posit that the relation between motivational constructs is bidirectional in the long run, some of these constructs might be strongly related to others in a given situation (Eccles & Wigfield, 2020; Marsh, 2023; Ryan & Deci, 2017). Understanding causal relations between various motivational constructs has been of a long interest in the field, especially due to its important implications for designing effective models of change for interventions. To explore such casual relations, scholars often chose to conduct experiments or longitudinal studies (Harackiewicz et al., 2012; Weidinger et al., 2017). These methods are certainly needed as manipulations, and time-lags are particularly useful for making causal inferences (Pearl, 2000). However, they can be time and resource intensive as well as subject to fat-handedness and omitted variable bias. Also, when only observational data can be gathered, alternative approaches are needed to examine causal relations. Therefore, in the present study, we aimed to triangulate plausible causal models in motivation research within educational psychology using directed acyclic graphs (DAGs) to understand the causal relationships among key constructs in Eccles and colleagues' Situated Expectancy–Value Theory (SEVT): expectancy for success, various subjective task values, and prior achievement.

Theoretical framework

We frame our work under SEVT (Eccles & Wigfield, 2020) to examine causal relationships among motivational constructs, as it is one of the most prominent theories (Wigfield et al., 2021). SEVT includes two central motivational beliefs for achievement-related choices and performance: expectancy for success and subjective task values. Expectancy for success is defined as students' beliefs about how well they will do on a task, and subjective task values is defined as students' desire to do the task. There are at least four facets of subjective task value: intrinsic value (i.e., interest), attainment value (i.e., identity), utility value (i.e., usefulness), and cost (i.e., what one has to lose or give up).

Previous findings using causally informative methodologies

To attempt to estimate the effects of SEVT constructs, scholars have either typically conducted randomized controlled trials (RCTs; Hulleman et al., 2010; Harackiewicz et al., 2012) or estimated cross-lagged panel models (CLPM; Dicke et al., 2018; Viljaranta et al., 2014; Weidinger et al., 2017) to infer causality. RCTs are a rigorous tool to investigate causal relations because observed and unobserved participant characteristics are balanced between treatment and control groups on expectation, reducing bias and attributing any differences in the outcome to the intervention (Hariton & Locascio, 2018). However, RCTs are often expensive and can be unethical as not all individuals receive the intervention (Faraoni & Schaefer, 2016; Sanson-Fisher et al., 2007). In the case of motivational interventions, the effect of the intervention is well-identified. However, localizing the effect to a specific motivational construct relies on the assumption that the intervention could not have affected outcomes of interest via other pathways, an assumption that may be violated in many psychological experiments. That is, RCTs likely suffer from fat-handedness, where multiple variables are changed at once, making it challenging to identify which variable caused the observed outcome (Eronen, 2020). For instance, interventions targeting utility value may impact both utility value and expectancy, as students are inquired to make connections between the learning content and their own lives, which requires some level of competence (see Hulleman et al., 2017; Rosenzweig et al., 2020). Thus, while such work has significant potential for practical applications, its theoretical implications require additional information for accurate interpretation.

Due to limited resources and other challenges described above, scholars have frequently used CLPM analyses of nonexperimental longitudinal data to examine causal relationships in longitudinal research (Campbell, 1963; Kenny & Harackiewicz, 1979). CLPMs estimate relations between two (or more) variables across time in which one variable from the first time point is investigated with regard to a second variable measured from the second time point and vice versa (Selig & Little, 2012). Unfortunately, CLPMs may suffer from omitted variable bias, resulting in biased estimates of causal effects (e.g., Berry and Willoughby, 2017; Hamaker et al., 2015). For example, studies using cross-lagged panel modeling have examined the co-development of a limited number of expectancy and subjective task value beliefs at once (e.g., Arens et al., 2019; Spinath & Steinmayr, 2008; Weidinger et al., 2019). However, the choice of SEVTs and control measures varies between studies, leading to incomplete empirical evidence and potentially different conclusions about the relationship between variables. For instance, a study controlling for between-class differences using a multi-level approach (Marsh et al., 2005) may produce different results than a study controlling only for prior intrinsic values (Spinath & Steinmayr, 2008). Further, in previous CLPMs in this area that statistically control for prior achievement, authors tend to find smaller cross-lagged paths than in studies that do not, consistent with the possibility that omitted variable bias may drive some of these associations (see Table S1 in Supplemental materials for examples). Therefore, caution should be taken when comparing results from models with different control measures to avoid incorrect interpretations of path coefficients.

Using psychometric networks and DAGs can bring several unique strengths that complement the contributions of RCTs and CLPMs. First, this method allows researchers to use cross-sectional data to reason about causal associations at a low cost. Although the hypothesized causal models need further testing, we do not have to manually fit and test all possible 59,049 structural equation models when using networks and DAGs with the Inferred-Causation Algorithm[1]. Second, unlike RCTs and CLPMs, which frequently focus on one or two variables at a time, networks/DAGs can examine causal relationships for a group of variables at once. For example, previous CLPM analyses among expectancy for success, subjective task values, and prior achievement tend to estimate causal relations among few—often just two constructs at a time, which makes omitted variables bias a likely threat to causal interpretations of estimates. However, psychometric network analyses allow researchers to holistically view the causal relations among these SEVT constructs (Tang et al., 2022). With the above-mentioned strengths, the current study, therefore, aimed to apply this novel approach to examine the causal relationships among expectancy for success, subjective task values, and achievement.

The present study

In the present study, we developed possible causal models based on SEVT constructs using networks/DAGs, compared the findings to prior findings that used different methods, and addressed the lack of formalized hypotheses using data and theory. We apply this approach to three different datasets of ninth graders in the mathematics domain to gauge the replicability of our findings due to concerns about the non-replicability of network models.[2] We chose the school subject of mathematics because it has been widely studied (e.g., Li et al., 2021; OECD, 2021; Stevenson & Stigler, 1994), particularly within SEVT (e.g., Eccles, 1994; Simpkins et al., 2006). In addition, understanding how we can improve our educational systems by leveraging the science of academic motivation has important implications for students' subsequent STEM opportunities, particularly during adolescence (Eccles, 1994; Simpkins et al., 2006).

Leveraging our three different datasets, we examined the following research questions (RQs):

RQ1. Is the magnitude of associations between SEVT constructs in the networks consistent across the three datasets?

We hypothesized consistent magnitudes of associations between SEVT constructs in the networks across datasets. Based on prior correlational findings, we expected: expectancy and intrinsic value to have the strongest association in the SEVT network (Eccles et al., 1983; Gaspard et al., 2018; Guo et al., 2016; Wigfield et al., 2009); expectancy to have weaker associations with utility value compared to intrinsic value and attainment value in the SEVT network (Gaspard et al., 2015a, b; Gråstén, 2016); and intrinsic value and attainment value to have a stronger association than intrinsic value and utility value in the SEVT network (Gaspard et al., 2015a, b; Conley, 2012).

RQ2. (a) What do the hypothesized directional paths look like between expectancy, subjective task values, and prior achievement using networks to create DAGs?; and (b) Are there consistent hypothesized directional paths in the three DAGs across datasets?

(a) Based on prior CLPM findings, we expected: achievement to predict expectancy (e.g., Viljaranta et al., 2014), intrinsic value (e.g., Garon-Carrier et al., 2016), utility value (e.g., Weidinger et al., 2019), and cost (e.g., Marsh et al., 2016); attainment value to predict achievement (e.g., Bonitto, 2020); intrinsic value to predict expectancy (e.g., Xu, 2018); expectancy to predict attainment value (e.g., Bonitto, 2020); and intrinsic to predict attainment value (e.g., Arens et al., 2019). Other interrelations between constructs were exploratory, as we did not have clear evidence from the literature to make directionality predictions.
(b) This research question was more exploratory. Despite differences in sampling variation, sample size, measures, and study design, we expected that there would be some consistent patterns within the DAGs, given the intended level of generalizability of SEVT.

Method

Datasets

Our main analyses used the California Achievement Motivation Project (CAMP) dataset (for more information, refer to Supplemental materials; Safavian & Conley, 2016; Safavian,

[1] This number was calculated by a combination of two out of five variables with three scenarios of directional path or 3^{10}, which equals 59,049 with no path or two directional paths. Note that some of these models are not identified for estimation.

[2] We chose ninth graders to create the final three DAGs for each dataset because this grade was consistently measured across all datasets.

2019). The original study design surveyed students' mathematics motivations using a cross-sequential design across middle to high schools located in the U.S. within a predominantly urban low-income area where most of the residents were foreign-born between 2004 and 2006. In addition, 39% of the school district was from Latin and South America. The study used one wave of students' ninth grade survey data ($n = 1{,}540$) collected in May 2005.

Datasets for replicability check

Two other datasets were examined to check the replicability of our findings across samples and measures (with replicability being defined as finding similar results with similar methods in a new dataset and a new context, such as a new country; see Schloss, 2018; Whitaker, 2017). Both datasets surveyed students' mathematics expectancy, subjective task values, and prior achievement in Germany. One of the datasets called the Motivation in Mathematics (MoMa) included 1,867 students in ninth grade (for more information, refer to Supplemental materials; Gaspard et al., 2015a, b). The other dataset called the Assessment of Students' Task Values in Secondary School (IF), included 103 students in ninth grade (for more information, refer to Supplemental materials; Gaspard et al., 2017).

Measures

Expectancy for success[3], subjective task values, and prior achievement were measured in all datasets. For this study, we focused on intrinsic, utility, and attainment value as subcomponents of subjective task value, but not cost, due to the lack of comparable items for the distinct facets of cost across datasets for our replicability check. Cost only had two items consistently across waves for the CAMP dataset, and the items did not match across different datasets, so they were excluded. We selected items with the purpose of having comparable items across datasets (see Table S2 to S4 for details about the cross-validation datasets measures).

CAMP measures

Expectancy for success Math expectancy for success was measured with three items (e.g., "How certain are you that you can learn everything taught in math?"; Midgley et al., 2000), each with a 5-point Likert scale ranging from 1 (*not at all X*) to 5 (*very X*) as the response format. All items were validated and came from the Academic Efficacy Scale (Midgley et al., 2000). The internal consistency of this scale was good ($\alpha = 0.83$).

Subjective task values Math subjective task values were measured using a 5-point Likert scale ranging from 1 (*not at all true of me*) to 5 (*very true for me*). Intrinsic value was measured by five items (e.g., "I enjoy the subject of math?"; Eccles and Wigfield, 1995; Conley, 2012). Utility value was measured by three items, tapping utility for future life and utility for daily life facets (e.g., "Math helps me in my daily life outside of school"; Eccles & Wigfield, 1995; Conley, 2012). Attainment value was measured by two items (e.g., "Thinking mathematically is an important part of who I am"; Eccles & Wigfield, 1995; Conley, 2012). All subjective task value items were derived from Eccles and Wigfields' (1995) work and were validated by Conley's (2012) work. The internal consistency of these scales was acceptable ($\alpha = 0.93$, 0.73, and 0.73, respectively).

Prior achievement Math prior achievement was operationalized using performance on the California Standards Tests (CSTs; California Department of Education, n.d.)—a statewide standardized math exam. CSTs scaled scores were obtained from school district records and ranged from 150 to 600, where higher numbers reflect higher scores.

Data analysis

Based on Waldorp (2020), we created networks to DAGs using the CAMP dataset with the following steps: (1) conducting confirmatory factor analysis (CFA) to obtain factor scores for expectancy for success, intrinsic, utility, and attainment value construct; (2) estimating undirected networks using the Gaussian Graphical Model (GGM; Lauritzen, 1996; Epskamp et al., 2018b); and (3) creating directed networks or DAGs using the Inferred-Causation Algorithm (ICA; Pearl, 1988; Pearl & Verma, 1991; see Table S5 in supplemental materials for a graphical overview of the analytical strategy). For more details on each step, refer to our Supplemental materials.

Obtaining factor scores To create a network for our DAGs, we needed to determine how the variables (or nodes, using network terminology) would be represented. Instead of using individual items, we chose to simplify the network at the construct level because different nodes representing the same psychological construct can affect which node is the most central or related to all other nodes in the network if item-level network analysis is conducted (Fried & Cramer, 2017). So, we conducted a CFA to obtain factor scores for each latent construct: expectancy for success, intrinsic value, utility value, and attainment value (Fig. 1).

[3] Academic self-efficacy was operationally defined as an expectancy for success measure, in order to stay consistent with terms used in SEVT (e.g., Guo et al., 2016; Simpkins et al., 2012; Wang & Eccles, 2013).

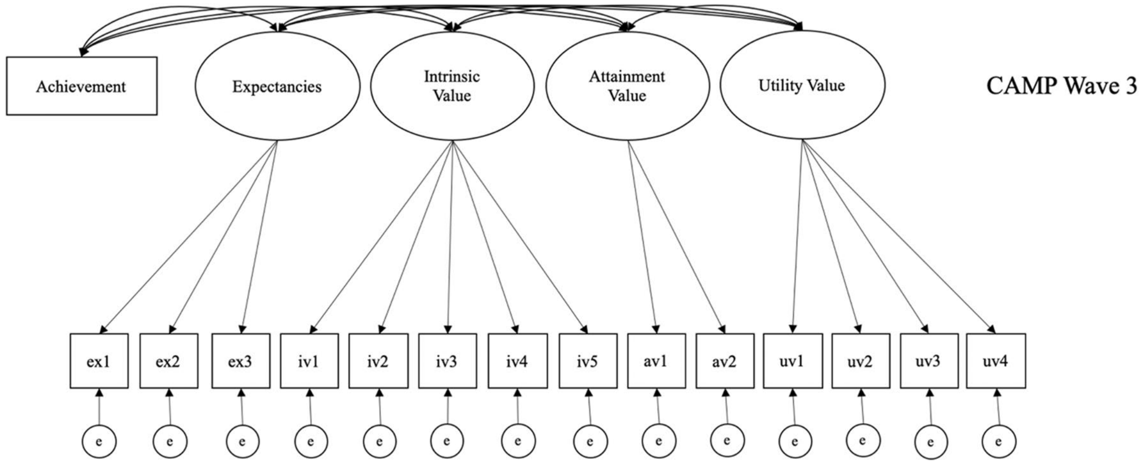


Fig. 1 Grade 9 factor model

Network estimation Using the factor scores along with achievement, undirected networks were estimated using the GGM (Lauritzen, 1996; Epskamp et al., 2018b). The connections between SEVT variables are represented as partial correlation coefficients. Positive and negative associations were indicated by green and red lines, respectively, with thicker lines representing stronger associations. To obtain a clear visual graph, the graphical Least Absolute Shrinkage and Selection Operator (LASSO) combined with the Extended Bayesian Information Criterion (EBIC) were used (Epskamp et al., 2018a, b).

Creating DAGs To create DAGs, a causal search algorithm called Inferred-Causation Algorithm (ICA; Pearl, 1988; Pearl and Verma, 1991) was conducted. This involved comparing zero-order correlation-based networks with partial correlation-based networks to identify plausible causal structures. In particular, causal inferences were made by examining colliders and confounders (see Edwards, 2000; Pearl, 2000; Rohrer, 2018 for more information).

Comparison of findings across datasets

To boost the credibility of our findings, different datasets were used in this study. As networks can vary with sample variation, sample size, measures, and study design, we wanted to ensure that at least some of our findings are replicable across samples before encouraging others to conduct further tests on them.

Comparing correlations and partial correlations Before creating possible DAG models, we investigated whether there were some replicable patterns between different datasets. If SEVT theory explains the process of achievement motivation observed, then the correlations and partial correlations for both datasets should be similar (see Supplemental materials for further description). To statistically compare the correlations and partial correlations outputs across datasets, we calculated and graphed the difference between correlations and partial correlations (i.e., correlation - partial correlation) and the ratio between correlations and partial correlations (i.e., partial correlation / correlation).

Comparing networks To test our hypotheses about finding replicability or consistency using networks, we compared the undirected networks between datasets by examining the visual graphs and edge strength outputs. We ensured that the nodes were graphically placed in similar positions in the network for visual comparison.

Comparing DAGs We compared DAGs across datasets by examining the visual directionality graphs created using the ICA.

Statistical software

All analyses were performed utilizing R (version 3.5.2; R Core Team, 2018). R-packages included: *qgraph* for network visualization and analysis (version 1.6.3; Epskamp et al., 2012) and *SIN* (version 0.6; Drton, 2013) as well as *mgm* (version 1.2-9; Haslbeck, 2020) for DAG analyses. For more details on these R-packages, see Supplemental materials.

Results

CAMP SEVT network estimation

Correlations and partial correlations among SEVT constructs are numerically presented in Table 1. Consistent

Table 1 Grade 9 correlations for factor scores psychometric network

Variable	(1)	(2)	(3)	(4)	(5)
CAMP					
1. Expectancies	–	0.35	0.34	-0.13	0.21
2. Intrinsic Value	0.72	–	0.31	0.04	0.05
3. Attainment Value	0.71	0.77	–	0.76	0.00
4. Utility Value	0.56	0.67	0.88	–	-0.05
5. Achievement	0.33	0.26	0.20	0.13	–
IF					
1. Expectancies	–	0.68	-0.24	0.39	0.29
2. Intrinsic Value	0.93	–	0.79	-0.59	-0.02
3. Attainment Value	0.86	0.94	–	0.77	-0.02
4. Utility Value	0.73	0.66	0.84	–	0.00
5. Achievement	0.58	0.50	0.46	0.40	–
MoMa					
1. Expectancies	–	0.60	0.00	-0.12	0.37
2. Intrinsic Value	0.85	–	0.61	0.00	-0.02
3. Attainment Value	0.69	0.87	–	0.50	-0.01
4. Utility Value	0.39	0.56	0.71	–	0.00
5. Achievement	0.57	0.44	0.37	0.22	–

Note. Correlations are represented below the diagonal; partial correlations are represented above the diagonal

with our hypothesis, in ninth grade for CAMP, we found that expectancy for success had the strongest tie with intrinsic value in the SEVT network ($r = .35$). We also found that expectancy for success had the weakest tie with utility value ($r = -.13$) compared to intrinsic value ($r = .35$) and attainment value ($r = .34$) in CAMP. Additionally, we found support for our third hypothesis where intrinsic value and attainment value ($r = .31$) had a stronger tie than intrinsic value and utility value ($r = .04$) in CAMP. The pattern of partial correlations is depicted as a network in Fig. 2. In the network, the thickness of the line indicates the strength between two nodes, while the color of the line indicates directionality (i.e., red = negative association vs. green = positive association). The line that connects two nodes (also known as an edge) represents partial correlations, which are numerically shown in Table 1 above the diagonal. Interestingly, there were also negative associations in our network, perhaps due to omitted variables or colliders, as all hypothesized effects are positive. We discuss this further in the "Discussion" section.

CAMP SEVT DAGs

First, using the SIN algorithm, the edges below the *p*-value of 0.10 were considered significant and important (Drton & Perlman, 2004, 2008; refer to Fig. 3).[4] Thus, the following seven edges[5] were considered to be significant: expectancy for success and intrinsic value, expectancy for success and attainment value, expectancy for success and utility value, expectancy for success and prior achievement, intrinsic value and attainment value, intrinsic value and prior achievement, and attainment value and utility value.

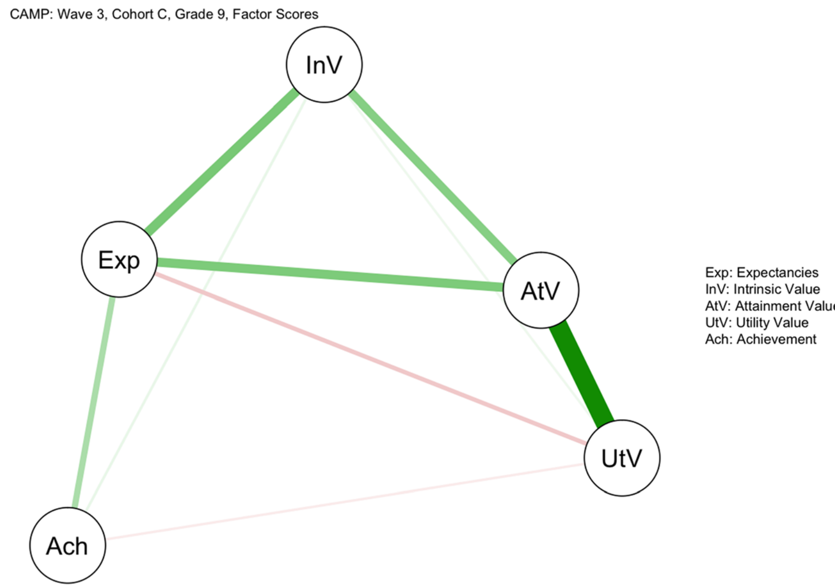
Next, we used the ICA (Pearl, 1988; Pearl & Verma, 1991) to reason from an undirected graph (i.e., a graph with no directionality) to a directed graph (i.e., a graph with directionality). We started looking at the relations between all five variables from our psychometric network because we wanted to find the set of nodes, S_{ab} that explains away the correlation between *a* and *b*. Namely, we wanted to find if any of the edges could be "empty" or "gone" after controlling for all other variables in the network—also known as the skeleton. Similar to the SIN algorithm results, we found the same edges present except for the association between utility value and prior achievement (refer to Fig. 4). We decided to work with the skeleton graph from the ICA because we aimed for a simple causal model.

To start drawing directionality for the DAG, we compared the correlation network to the partial correlation network (refer to Fig. 5). We looked to see which edges either "disappeared" or "appeared" between the two networks, referring to whether or not there is a zero-order correlation. Results found the following changes between the correlation and partial correlation network: (1) the connection between intrinsic value and utility value was gone in the partial correlation network; and (2) the connection between intrinsic value and prior achievement was gone in the partial correlation network.

[4] To determine statistical significance, a *p*-value of 0.05 is typically the adopted rule of thumb. In this study, we followed the Significance, Intermediate, and Non-significance (SIN) approach (Drton & Perlman, 2004, 2008), which proposes to use 0.10 as the threshold for the significance. There are several reasons for using this *p*-value as a criterion. First, the *p*-values reported in this study have been corrected for multiple comparisons. The *p*-value adjustments will increase the size of *p*-values, making it harder to reach the 0.05 threshold. Second, 0.05 is known as an arbitrary threshold to define statistical significance (see commentaries by Amrhein et al., 2019; Wasserstein et al., 2019). The threshold defining statistical significance should be chosen on the basis of subjective and contextual conditions (Drton & Perlman, 2004). In the context of psychometric network analyses, 0.10 is a more realistic threshold to determine the significance of the edges because many variables have been examined simultaneously, *p*-value adjustments have been applied, and because some paths in the network analysis are likely downwardly biased, because colliders or mediators are controlled. Please note that we only choose this threshold to find meaningful DAGs, all our significant DAG results are below 0.05 (see Fig. 3, S2, and S7).

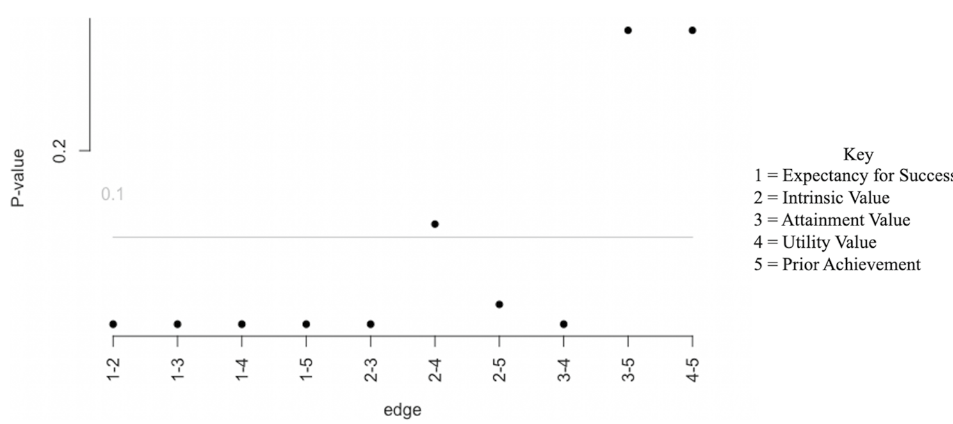
[5] Exp = expectancy for success; Int = intrinsic interest value; AtV = attainment value; UtV = utility value; and Ach = prior achievement.

Fig. 2 CAMP Grade 9 psychometric network models



Note. Psychometric networks are constructed using partial correlation coefficients where the association between two nodes controls for all other information. In the network, the thickness of the line indicates the strength between two nodes, while the color of the line indicates directionality (i.e., red = negative association vs. green = positive association). Networks used Extended Bayesian Information Criterion (EBIC) LASSO regularization to limit the number of spurious edges

Fig. 3 CAMP Grade 9 significant partial correlations



Note. Significant, Intermediate, Non-significant (SIN) algorithm tested each partial correlation against zero. The x-axis indicates all possible connections; the y-axis indicates p-values. All below 0.1 are significant

Before testing these relationships, we pointed prior achievement to expectancy for success because of temporal relations, and prior achievement has been found to influence expectancy (Eccles, 2009; Viljaranta et al., 2014). Moreover, prior achievement was added to both the network and DAG because prior achievement is a major potential confounder that could influence relations among SEVT constructs, although it is not always included in previous studies of dynamics among SEVT constructs (Liu, 2016; Pinxten et al., 2014).

Then we selected three variables at a time. To draw the directionality for the relationship between intrinsic value and utility value, we chose various possible combinations of variables. For example, we examined the relations between

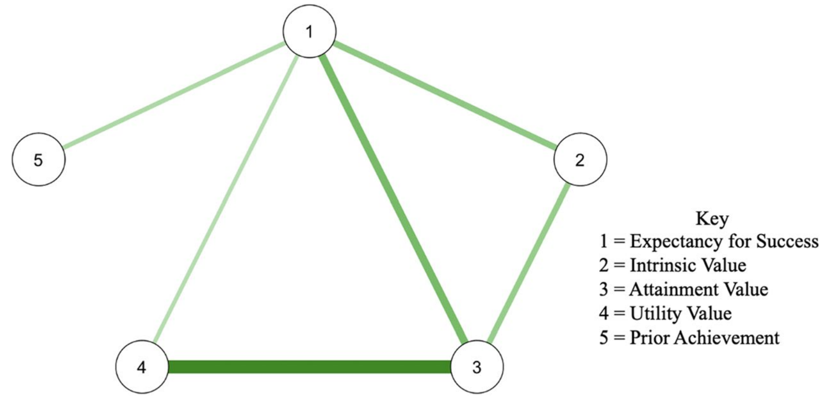


Fig. 4 CAMP Grade 9 inferred causation algorithm all five variables

Note. Network examines the relationship between all five SEVT variables using the Inferred Causation Algorithm. This network displays which variables are connected to each other. The presence of an edge (i.e., line) means that there is a connection between two variables, while the absence of an edge means that there is no connection between two variables

(1) intrinsic value, utility value, and expectancy for success (2Int-4UtV-1Exp[6]), (2) intrinsic value, utility value, and attainment value (2Int-4UtV-3AtV), and (3) intrinsic value, utility value, and prior achievement (2Int-4UtV-5Ach; refer to Fig. 7). We found that the connection between intrinsic value and utility value (as shown in the partial correlation network) disappeared when attainment value was present. Therefore, we inferred that there could be a collider effect where both intrinsic value and utility value pointed to attainment value (2Int → 3AtV ← 4UtV). There was no need to examine the relations between four variables (e.g., 2Int-4UtV-1Exp-3AtV or 2Int-4UtV-1Exp-5Ach or 2Int-4UtV-3AtV-5Ach) because we already directed the lines.

To draw the directionality for the relationship between intrinsic value and prior achievement, we again chose various possible combinations of variables. For example, we examined the relations between (1) intrinsic value, prior achievement, and expectancy for success (2Int-5Ach-1Exp), (2) intrinsic value, prior achievement, and attainment value (2Int-5Ach-3AtV), and (3) intrinsic value, prior achievement, and utility value (2Int-5Ach-4UtV; refer to Fig. 6). We found that the link between intrinsic value and prior achievement (as shown in the partial correlation network) disappeared when expectancy for success was present. Therefore, we inferred that there could be a collider effect where both intrinsic value and prior achievement pointed to expectancy for success (2Int → 1Exp ← 5Ach). There was no need to examine the relations between four variables (e.g., 2Int-5Ach-1Exp-3AtV or 2Int-5Ach-1Exp-4UtV or 2Int-5Ach-3AtV-4UtV) because we already directed the lines.

Now that the directionality of effects between interest value and prior achievement was depicted, we examined the partial correlation network to seek further possible directed lines. We found that the association between expectancy for success and utility value as well as expectancy for success and attainment value was still missing (or non-directed) in our DAG. So, we started with the association between expectancy for success and utility value. We first tried pointing utility value to expectancy for success. However, this option was not possible because there should have been a zero-order correlation between utility value and prior achievement due to a common effect. So, the only other option was to point expectancy for success to utility value.

Next, we examined the association between expectancy for success and attainment value. We first tried pointing attainment value to expectancy for success. But this was not a possible option for the DAG because then there was a cyclical relationship among attainment value, expectancy for success, and utility value (i.e., 3AtV → 1Exp → 4UtV → 3AtV), which meant the graph was not acyclic anymore. So, the only other option was to point expectancy for success to attainment value. As a result, we ended up with Fig. 7 using the CAMP dataset. The following lines were directed: both prior achievement and intrinsic value pointed to expectancy for success, both intrinsic value and utility value pointed to attainment value, and expectancy for success both pointed to utility value and attainment value.

[6] The numbers in front of the SEVT construct abbreviations are used to match the same numbers used in the key from prior figures, in order to stay consistent.

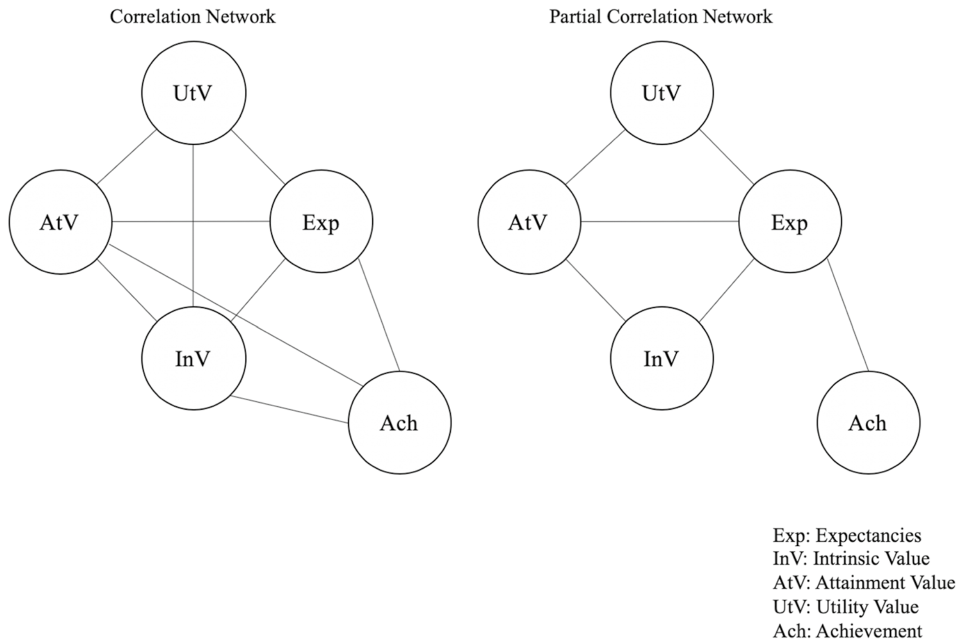


Fig. 5 CAMP Grade 9 correlation and partial correlation network

Note. Models are in the following order: correlation network and partial correlation network. The thickness of the lines is not represented here because we are just comparing the presence or absence of an edge between the correlation and partial correlation network

Replicability

We found mostly consistent results across different samples and both countries (refer to Figs. S1 to S12 in Supplemental materials).

Comparing correlations and partial correlations

We found similar correlation and partial correlation patterns across different datasets and countries. Results showed that the associations of SEVT constructs between datasets are not spurious when examining the correlation and partial correlation difference and ratio graph (see Fig. S13 in Supplemental materials). In the correlation and partial correlation difference graph, we found points on the left side of the axis, meaning there was no large difference between the correlations in various datasets. In the correlation and partial correlation ratio graph, we found points on the right side of the axis, meaning there was no large difference between the correlations in various datasets. Additionally, we found that the partial correlations decrease and go in the opposite direction, as predicted by theory when compared to the correlations.

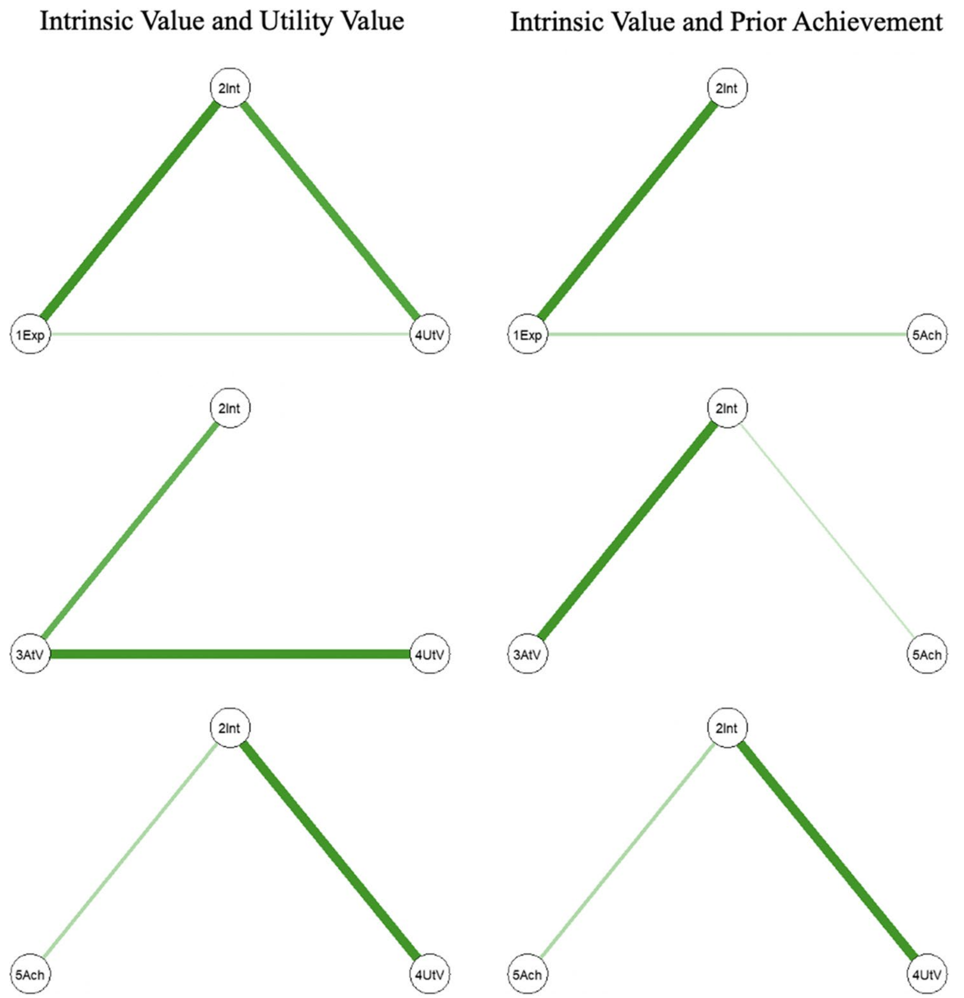
Comparing networks

Across the three datasets, only expectancy is correlated with the achievement after partialling out all other variables (r = .21; 0.29; and 0.37), indicating that previous achievement is likely closer to the expectancy (see Table 1). Further, expectancy is highly associated with intrinsic value after partialling out common variances (r = .35; 0.68; 0.60), and intrinsic value is positively associated with only attainment value in the partial correlation matrix across three samples (r = .31; 0.79; 0.61; see Table 1). Finally, the associations of variables with utility value vary the most across datasets and countries, indicating that the role of utility value would be expected to vary across context and culture. For example, we found that utility value had the weakest tie with expectancy for success in CAMP and MoMa, but not IF (see Table 1). For a visual representation of the networks, see Fig. S13 in Supplemental materials.

Comparing directed acyclic graph models

Three DAGs were formulated based on the various datasets using the data from ninth grade (see Fig. 8). We found similar DAG models when comparing the CAMP and MoMa data. For example, prior achievement pointed to expectancy for success, intrinsic value pointed to expectancy for success, and utility value pointed to attainment value. However, we see a different structure for IF, possibly due to the small sample size. In this case, both utility value and attainment value pointed to expectancy for success, and utility value pointed to intrinsic value. We expand upon the difference between the DAGs in the discussion.

Fig. 6 CAMP Grade 9 inferred causation algorithm three variables



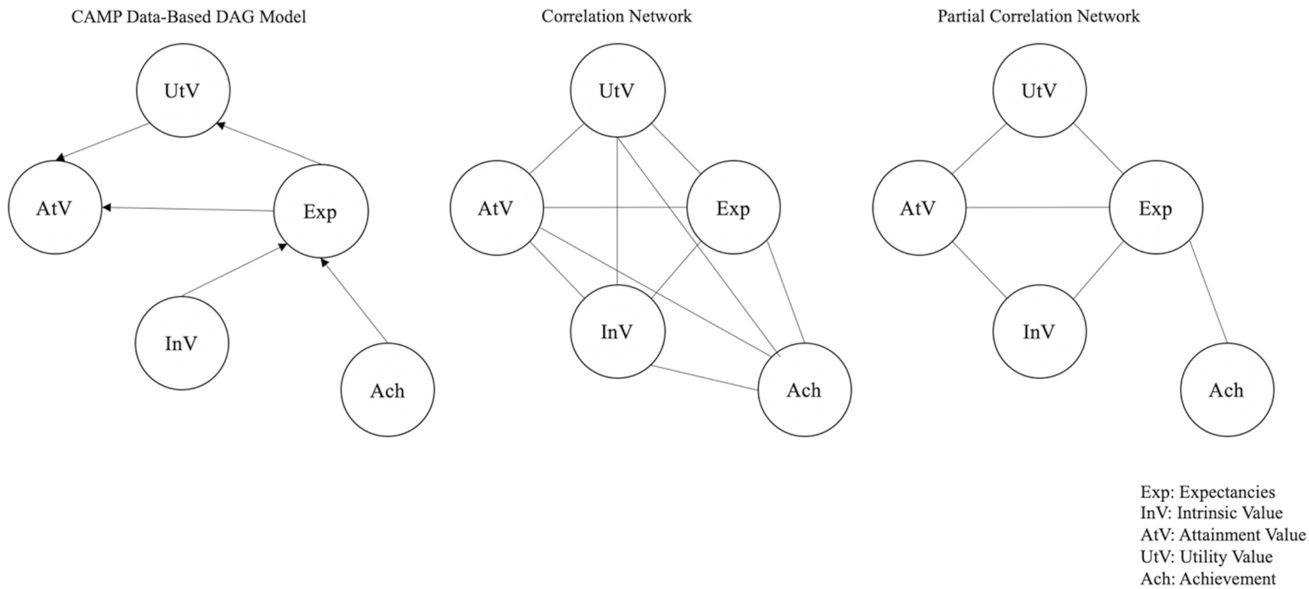
Note. Examining the relationship between three variables that include intrinsic value and utility value as well as intrinsic value and prior achievement using the Inferred Causation Algorithm. Exp = Expectancy for success; Int = Intrinsic value; AtV = Attainment value; UtV = Utility value; and Ach = Achievement. The numbers in front of the SEVT construct abbreviations are used to match the same numbers used in the key from prior figures, in order to stay consistent. There was no need to examine the relationship between four variables (e.g., 2Int-4UtV-1Exp-3AtV or 2Int-4UtV-1Exp-5Ach or 2Int-4UtV-3 AtV-5Ach) because we already directed the lines

Discussion

In the present study, we used psychometric network analysis and DAGs to generate testable hypotheses about the effects among SEVT constructs. We have formulated the hypotheses itself the focus of this study so that the hypotheses are the result rather than the starting point of this study. Starting only with the assumption that expectancy does not point to prior achievement, but prior achievement influences expectancy due to temporal relations, we identify some consistent patterns of similarities and differences in the associations and hypothesized relations across three different datasets.

RQ1. Psychometric networks of students' expectancy, subjective Task values, and prior achievement

Several findings further our understanding of the interplay of adolescents' expectancy and subjective task values in



Note. Models are in the following order: directed acyclic graph, correlation network, and partial correlation network

Fig. 7 CAMP Grade 9 directed acyclic graph model

mathematics. As hypothesized, we found consistent similarities in the strengths of the relations among the constructs across the three datasets and two countries. First, students' expectancy for success had the strongest tie with intrinsic value in all the datasets. This finding supports previous research highlighting the connection between competence and interest, particularly in contexts with high standards of excellence (e.g., engaging students in activities that provide

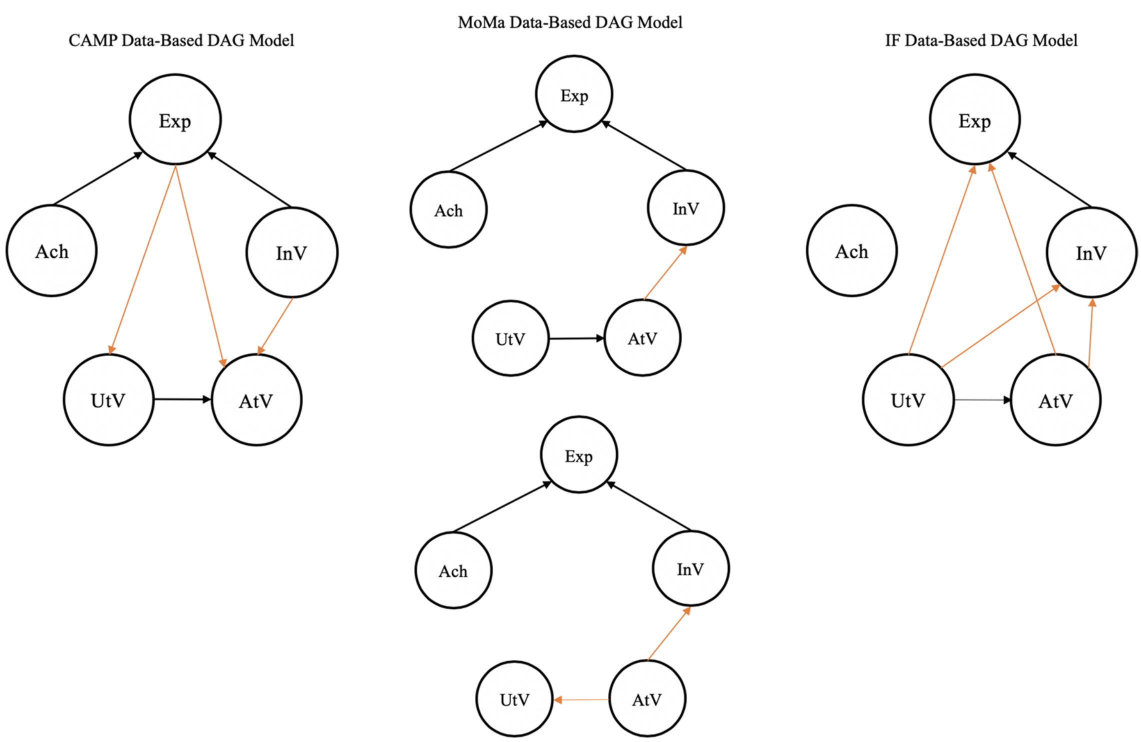


Fig. 8 Directed acyclic graph model across datasets. *Note*. Orange lines reflect paths that differ between the directed acyclic graphs

feedback; Bandura, 1986; Eccles, 2005; Wigfield, 1994). The second most consistent finding was the weakest edge between expectancy for success and utility value (particularly in the two largest data sets) compared to the other subjective task values in the SEVT network. In retrospect, the unexpectedly weaker linkages of expectancy to utility value perhaps can be explained in terms of broader perceptions of mathematics' usefulness. Many students likely know that mathematics is useful in many ways, yet still do not expect they can do well in it, thus reducing the edges between the two constructs. Third, we found that students' intrinsic and attainment values were more strongly correlated with each other, compared to intrinsic value and utility value, which aligned with our hypothesis. These findings are consistent with previous work on the relations of these components of task value in different studies, some of which used the same datasets as we did (Eccles & Wigfield, 1995; Gaspard et al., 2018; Guo et al., 2016; Trautwein et al., 2012). Eccles, Wigfield, and their colleagues have discussed how both intrinsic and attainment value reflect intrinsic/internal characteristics of the individual and how the activity relates to those perceptions, as compared to utility value which focuses more on what is considered to be extrinsic aspects for doing the task (e.g., "I need to take this math course in order to become a doctor"; Eccles, 2005; Eccles & Wigfield, 2020).

RQ2a. Hypothesized DAGs of students' expectancy, subjective Task values, and prior achievement

The most novel contribution of the results was the hypothesized directional models we developed with the help of DAGs. First, we found that intrinsic value was hypothesized to point to expectancy for success across datasets. This finding suggested that the more an individual enjoys a task, the more one believes they can do well. Perhaps, intrinsic value promotes engagement and involvement, which then promotes expectancy for success. The different DAGs showed that expectancy for success was a "collider" of prior achievement and intrinsic value – indicating that both prior achievement and intrinsic value cause expectancy for success, or vice versa. This result regarding the hypothesized directionality of the intrinsic value-expectancy for success relationship aligns with previous empirical research (Denner et al., 2019; Nuutila et al., 2018; Pinxten et al., 2014; Yoon, 1996).

Additionally, results showed that utility value was hypothesized to point to attainment value across datasets. Although SEVT does not predict the direction of the order, this finding is consistent with the developmental sequence implicit in the motivational orientation of Self-Determination Theory (SDT). Within SDT, Deci and Ryan's (1985) Organismic Integration Theory posits that external regulation precedes identified or integrated regulation. A potential interpretation for this finding consistent with SDT is that a more external self-belief like utility value causes one to eventually internalize a self-belief like attainment value.

RQ2b. Replicability of the Findings across datasets

There was much overlap across the datasets, but also some notable differences. For example, in the dataset used for main analyses, both intrinsic value and utility value were hypothesized to point to attainment value, yielding attainment value as a collider variable. In contrast, in the other two datasets used for replication purposes, utility value was hypothesized to point to attainment value which then points to intrinsic value, yielding a chain structure. One broad explanation for these discrepant findings is that students from the two datasets used for replication purposes come from more similar backgrounds. The main analyses presented in this paper come from a U.S. sample, whereas the other two datasets used for replication come from a German sample. The U.S. sample represents students from an urban, predominantly low-income immigrant background. Half of the students in this sample perform below average in the mathematics state standards assessments and are enrolled in remediation courses such as pre-algebra (Safavian, 2019). In contrast, the German sample represents students in academically tracked schools where mathematics is compulsory—not differentiated by level—so all the students in these schools learn algebra, geometry, and calculus in one comprehensive course (Gaspard et al., 2015a). It is theoretically plausible that the opportunity to develop an interest in mathematics for students in the U.S. comes from self-discovery (i.e., time reflecting on the emotional experiences of enjoyment) as they are not afforded the experiences that prompt them to think about their identity as a mathematics person. Whereas students in Germany may have been exposed to socialization experiences that prompt them to think about identity early on, in terms of being a mathematics person or mathematics being important to their sense of who they are. Yet, we interpret differences in hypothesized directionality across datasets with caution as our analysis is unable to differentiate contextual moderation from sampling variability. Therefore, in this study, we focused on examining the replication of our findings across different datasets instead of investigating cross-country differences because clear consistencies across different populations suggest that our approach can detect replicable patterns in correlation and partial correlation matrices within the SEVT literature.

We, however, encourage researchers to probe the potential influence of specific contextual differences more systematically across different countries in the future, especially because Eccles and Wigfield (2020) emphasize the "situated" nature of SEVT. For example, researchers should closely study how students across various countries develop expectancy and subjective task values as the kinds of mathematics curricula, classroom structure, teacher-student relations, school funding, and socioeconomic status can impact them differently. Understanding how contextual differences can shape students' expectancy and subjective task values has important implications for intervention design, considering that a one-size-fits-all intervention will be less likely to help such different populations.

Methodological implications

With this study, we hoped to pave a path for more causal thinking and formalized, testable hypotheses that should guide future research and serve as bridges of insight between the findings from decades of previous cross-sectional research and future studies hoping to test directed causal hypotheses. This undertaking is inspired by ongoing debates about the so-called theory crisis, in the context of which researchers have pointed out the lack of formalized theories and specific testable hypotheses (Eronen & Bringmann, 2021; Eronen & Romeijn, 2020; Oberauer & Lewandowsky, 2019). We hope to have contributed to the generation of such testable hypotheses. This study reacts to recent arguments addressing the lacking quality of research hypotheses and the lack of testable hypotheses (Borsboom et al., 2021). An avenue for further study is a longitudinal extension allowing for SEVT constructs to serve as both cause and effect of another in a dynamical system (causal cycle). While we believe in the potential usefulness of network models for an improved understanding of the relations among SEVT facets with another, future work should extend this method to longitudinal data, as Moeller et al. (2021) suggested. The network methods used in our study (DAG and ICA) are not suited to study such reciprocal causal feedback loops (causal cycles; Ryan et al., 2019). To avoid confusion among these different approaches of using networks to understand aspects of causality among SEVT facets, future studies need to distinguish between different aspects of causality and the appropriate network methods.

Limitations

There are limitations to the present research. First, all data were cross-sectional, and therefore no certain conclusions about causal directionality can be drawn. Our conclusions are rather possible causal directional path models that can be triangulated with current empirical research. In other words, our SEVT models present a simplified *depiction of hypothesized* causal relations among SEVT constructs. With the method employed (the ICA using psychometric networks and DAGs), we hoped to create hypotheses for plausible causal paths, but importantly, these hypotheses were formulated a posteriori after exploring the covariance structure and consequently need to be tested further in more deductive approaches with data allowing for such directed interpretations (e.g., experimental and/or longitudinal). Moreover, the differences in the DAGs across datasets could be due to sampling error as well as moderation by contextual and time factors. Since psychometric networks are sensitive to sample size, we worry that rather than contextual moderation, differences in the DAGs may be due to the small sample size, especially since one of the replication datasets with the smallest sample size yielded the most discrepant results (Epskamp et al., 2018a, b) and especially since this small dataset was smaller than the 250 participants that are required to obtain stable, trustworthy correlation effect sizes (Schönbrodt & Perugini, 2013). In addition, the causal influence of unobserved variables cannot be directly measured in our work, yet these omitted variables may have biased our estimates, such as yielding negative estimates. These limitations can be alleviated to the extent that future studies depend upon the kinds of variables measured and included in the analyses. Future analyses can make use of longitudinal data to test the robustness of results to various assumptions about the structure of unobserved confounding over time (e.g., Zyphur et al., 2020).

Future research directions

Our work is the first step to delineating the questions that future work can answer using a combination of longitudinal data and exogenous variation, such as a randomly assigned manipulation. Work testing the emergent DAGs within this study in different contexts would provide needed information about whether similar patterns arise among the SEVT constructs. Once (or if) DAGs are robust across similar samples and contexts, then differences in DAGs from different samples and contexts can be hypothesized and experimentally tested. The ability to explain these differences is important because individuals will always be different in many ways. These individual differences are especially important to understand as there is a prominent issue of "whiteness" in motivation research (DeCuir-Gunby & Schutz, 2017; Usher, 2018). Motivation is not a one-size-fits-all because our attitudes about "Can I do it?" and "Do I want to do it?" comes from reflections of what we internalized over time as a function of our experiences and the experiences of those around us. If we are to truly design effective interventions, we must know how these constructs work for the specific population we plan to conduct the intervention. We hope this work provides an additional tool that can serve the need to study motivation in racially and ethnically diverse populations to serve them better. We also suggest that researchers use network analyses to investigate the interplay of the three value components studied here along with perceived cost, given the recent interest in the cost construct in the field (see Barron & Hulleman, 2015; Eccles & Wigfield, 2020; Wigfield

& Eccles, 2020). Unfortunately, we could not include cost in all our networks because of the inconsistent use of cost items across datasets. Ultimately, examining how the interplay of individuals' expectancy and values influences performance and choice will be a key step forward and perhaps provide information that researchers developing new SEVT-based interventions will find particularly relevant or useful.

Conclusion

We have contributed to the assessment and further development of the SEVT model by assessing specific DAGs that can be viewed as the hypothesized processes that underlie the relation between motivation constructs and achievement. The specificity of the paths represented in the DAGs form provides specific testable predictions about the interplay between SEVT constructs. In sum, we found that prior achievement relates to expectancy, intrinsic value relates to expectancy, and utility relates to attainment value, which can be used to triangulate experimental and longitudinal work.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s12144-023-04871-z.

Acknowledgements We thank our research assistants, Yiqun Zhang and Mingmei (Fern) Zhang, for working on Table S1.

Funding This work was supported by the National Science Foundation Graduate Research Fellowship to Hye Rin Lee [award number DGE-1,839,285], the National Science Foundation to Stuart Karabenick and Martin Maehr (DLR-0335369) and Jacquelynne S. Eccles, Stuart Karabenick, and Nayssan S. Safavian (HRD-1,535,273), the German Research Foundation Grant (TR 553/7 – 1) to Ulrich Trautwein, Oliver Lüdtke, and Benjamin Nagengast, and the LEAD Graduate School & Research Network [GSC1028], which was funded within the framework of the Excellence Initiative of the German federal and state governments.

Data availability The data from the current study can be obtained from the corresponding author upon reasonable request.

Declarations

This study was approved by the Institutional Review Board.

Informed consent All participants in the study provided informed consent.

Conflict of interest The authors declare no conflict of interest.

References

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*(7748), 305–307. https://doi.org/10.1038/d41586-019-00857-9

Arens, A. K., Schmidt, I., & Preckel, F. (2019). Longitudinal relations among self-concept, intrinsic value, and attainment value across secondary school years in three academic domains. *Journal of Educational Psychology, 111*(4), 663–684. https://doi.org/10.1037/edu0000313

Bandura, A. (1986). Fearful expectations and avoidant actions as coeffects of perceived self-inefficacy. *American Psychologist, 41*(12), 1389–1391. https://doi.org/10.1037/0003-066X.41.12.1389

Barron, K. E., & Hulleman, C. S. (2015). Expectancy-value-cost model of motivation. *Psychology, 84*, 261–271.

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: rethinking a developmental workhorse. *Child Development, 88*(4), 1186–1206. https://doi.org/10.1111/cdev.12660

Bonitto, C. (2020). Comparing male and female student self-efficacy, identification, and achievement in mathematics: a cross-lagged panel analysis of causal effects [Doctoral Dissertation, University of Kansas]. https://kuscholarworks.ku.edu/bitstream/handle/1808/31849/Bonitto_ku_0099D_17307_DATA_1.pdf?sequence=1. Accessed 12/10/20.

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: a practical framework for building theories in psychology. *Perspectives on Psychological Science, 17*(4), 756–766. https://doi.org/10.1177/1745691620969647

California Department of Education. (n.d.). California standardized testing and reporting program. http://star.cde.ca.gov/star2005/aboutSTAR.asp. Accessed 12/10/20.

Campbell, D. T. (1963). From description to experimentation: Interpreting trends as quasi-experiments. In C. W. Harris (Ed.), *Problems in measuring change*. University of Wisconsin Press.

Conley, A. M. (2012). Patterns of motivation beliefs: combining achievement goal and expectancy-value perspectives. *Journal of Educational Psychology, 104*(1), 32–47. https://doi.org/10.1037/a0026042

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.

DeCuir-Gunby, J. T., & Schutz, P. A. (Eds.). (2017). *Race and ethnicity in the study of motivation in education*. Routledge.

Denner, J., Valdes, O., Dickson, D. J., & Laursen, B. (2019). Math interest and self-concept among latino/a students: reciprocal influences across the transition to middle school. *Journal of Adolescence, 75*(1), 22–36. https://doi.org/10.1016/j.adolescence.2019.06.015

Dicke, T., Marsh, H. W., Parker, P. D., Pekrun, R., Guo, J., & Televantou, I. (2018). Effects of school-average achievement on individual self-concept and achievement: unmasking phantom effects masquerading as true compositional effects. *Journal of Educational Psychology, 110*(8), 1112–1126. https://doi.org/10.1037/edu0000259

Drton, M. (2013). SIN: A SINful approach to selection of gaussian graphical markov models. R package version 0.6. https://CRAN.R-project.org/package=SIN. Accessed 12/10/20.

Drton, M., & Perlman, M. D. (2004). *A SINful approach to Gaussian graphical model selection*. https://arxiv.org/abs/math/0508267v1

Drton, M., & Perlman, M. D. (2008). A SINful approach to Gaussian graphical model selection. *Journal of Statistical Planning and Inference, 138*(4), 1179–1200. https://doi.org/10.1016/j.jspi.2007.05.035

Eccles, J. S. (1994). Understanding women's educational and occupational choices: applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly, 18*, 585–609.

Eccles, J. S. (2005). Subjective task value and the eccles et al. model of achievement-related choices. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 105–121). Guilford Publications.

Eccles, J. S. (2009). Who am I and what am I going to do with my life? Personal and collective identities as motivators of action. *Educational Psychologist, 44*, 78–89. https://doi.org/10.1080/00461520902832368

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: the structure of adolescents' achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21*(3), 215–225. https://doi.org/10.1177/0146167295213003

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: a developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75–146). W. H. Freeman.

Edwards, D. (2000). *Introduction to graphical modelling* (2nd ed.). Springer Verlag.

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*, 1–18. https://doi.org/10.18637/jss.v048.i04

Epskamp, S., Borsboom, D., & Fried, E. I. (2018a). Estimating psychological networks and their accuracy: a tutorial paper. *Behavior Research Methods, 50*, 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018b). The gaussian graphical model in cross-sectional and time-series data. *Multivariate Behavioral Research, 53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: how to move forward. *Perspectives on Psychological Science, 16*(4), 779–788. https://doi.org/10.1177/1745691620970586

Eronen, M. I., & Romeijn, J. W. (2020). Philosophy of science and the formalization of psychological theory. *Theory and Psychology, 30*(6), 786–799. https://doi.org/10.1177/0959354320969876

Faraoni, D., & Schaefer, S. T. (2016). Randomized controlled trials vs. observational studies: why not just live together? *BMC Anesthesiology, 16*(1), 1–4. https://doi.org/10.1186/s12871-016-0265-3

Fried, E. I., & Cramer, A. O. (2017). Moving forward: challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science, 12*, 999–1020. https://doi.org/10.1177/1745691617705892

Garon-Carrier, G., Boivin, M., Guay, F., Kovas, Y., Dionne, G., Lemelin, J. P., Séguin, J. R., Vitaro, F., & Tremblay, R. E. (2016). Intrinsic motivation and achievement in mathematics in elementary school: a longitudinal investigation of their association. *Child Development, 87*(1), 165–175. https://doi.org/10.1111/cdev.12458

Gaspard, H., Dicke, A. L., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015a). Fostering adolescents' value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology, 51*(9), 1226–1240. https://doi.org/10.1037/dev0000028

Gaspard, H., Dicke, A. L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015b). More value through greater differentiation: gender differences in value beliefs about math. *Journal of Educational Psychology, 107*, 663–677. https://doi.org/10.1037/edu0000003

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology, 48*, 67–84. https://doi.org/10.1016/j.cedpsych.2016.09.003

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: how academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology, 52*, 1–14. https://doi.org/10.1016/j.cedpsych.2017.10.003

Gråstén, A. (2016). Children's expectancy beliefs and subjective task values through two years of school-based program and associated links to physical education enjoyment and physical activity. *Journal of Sport and Health Science, 5*(4), 500–508. https://doi.org/10.1016/j.jshs.2015.12.005

Guo, J., Nagengast, B., Marsh, H. W., Kelava, A., Gaspard, H., Brandt, H., ... & Brisson, B. (2016). Probing the unique contributions of self-concept, task values, and their interactions using multiple value facets and multiple academic outcomes. *AERA Open, 2*(1), 1–20.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102–116. https://doi.org/10.1037/a0038889

Harackiewicz, J. M., Rozek, C. S., Hulleman, C. S., & Hyde, J. S. (2012). Helping parents to motivate adolescents in mathematics and science: an experimental test of a utility-value intervention. *Psychological Science, 23*(8), 899–906. https://doi.org/10.1177/0956797611435530

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics and Gynaecology, 125*(13), 1716. https://doi.org/10.1111/1471-0528.15199

Haslbeck, J. M. B., & Waldorp, L. J. (2020). Mgm: estimating time-varying mixed graphical models in high-dimensional data. *Journal of Statistical Software, 93*, 1–46.

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*(4), 880–895. https://doi.org/10.1037/a0019506

Hulleman, C. S., Kosovich, J. J., Barron, K. E., & Daniel, D. B. (2017). Making connections: replicating and extending the utility value intervention in the classroom. *Journal of Educational Psychology, 109*(3), 387–404. https://doi.org/10.1037/edu0000146

Kenny, D. A., & Harackiewicz, J. M. (1979). Cross-lagged panel correlation: practice and promise. *Journal of Applied Psychology, 64*(4), 372–379. https://doi.org/10.1037/0021-9010.64.4.372

Lauritzen, S. L. (1996). *Graphical models*. Clarendon Press.

Li, Q., Cho, H., Cosso, J., & Maeda, Y. (2021). Relations between students' mathematics anxiety and motivation to learn mathematics: a Meta-analysis. *Educational Psychology Review, 33*, 1017–1049. https://doi.org/10.1007/s10648-020-09589-z

Liu, Y. (2016). *Effects of intrinsic and extrinsic motivation on academic achievement: Are they antagonistic, additive or multiplicative*? (Publication No. 1943010351) [Doctoral dissertation, The Chinese University of Hong Kong]. ProQuest Dissertations & Theses Global.

Marsh, H. W. (2023). Extending the reciprocal effects model of math self-concept and achievement: long-term implications for end-of-high-school, age-26 outcomes, and long-term expectations. *Journal of Educational Psychology, 115*(2), 193–211. https://doi.org/10.1037/edu0000750

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: reciprocal effect models of causal ordering. *Child Development, 76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Marsh, H. W., Pekrun, R., Lichtenfeld, S., Guo, J., Arens, A. K., & Murayama, K. (2016). Breaking the double-edged sword of effort/trying hard: developmental equilibrium and longitudinal relations among effort, achievement, and academic self-concept. *Developmental Psychology, 52*(8), 1273–1290. https://doi.org/10.1037/dev0000146

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E. M., Anderman, L., Freeman, K. E., ... Urdan, T. (2000). *Manual for the Patterns* *of Adaptive Learning Scales (PALS)*. Ann Arbor: University of Michigan.

Moeller, J., Viljaranta, J., Tolvanen, A., Kracke, B., & Dietrich, J. (2021). Introducing the DYNAMICS Framework of moment-to-moment development in achievement motivation. Pre-print PsyArXiv: https://psyarxiv.com/6brsa/

Nuutila, K., Tuominen, H., Tapola, A., Vainikainen, M. P., & Niemivirta, M. (2018). Consistency, longitudinal stability, and predictions of elementary school students' task interest, success expectancy, and performance in mathematics. *Learning and Instruction, 56*, 73–83. https://doi.org/10.1016/j.learninstruc.2018.04.003

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review, 26*, 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

OECD (2021). Mathematics performance (PISA) (indicator). https://doi.org/10.1787/04711c74-en Accessed on 28 Sept 2021.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann Publishers Inc.

Pearl, J. (2000). *Causality*. Cambridge University Press.

Pearl, J., & Verma, T. (1991). A theory of inferred causation. In *Proceedings of the Second International Conference of Representation and Reasoning*. San Francisco: Morgan Kaufmann.

Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology*, *84*(1), 152–174. https://doi.org/10.1111/bjep.12028

R Core Team (2018) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org. Accessed 12/10/20.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science, 1*(1), 27–42.

Rosenzweig, E. Q., Wigfield, A., & Hulleman, C. S. (2020). More useful or not so bad? Examining the effects of utility value and cost reduction interventions in college physics. *Journal of Educational Psychology, 112*(1), 166–182. https://doi.org/10.1037/edu0000370

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: basic psychological needs in motivation, development, and wellness*. Guilford Publications.

Ryan, O., Bringmann, L. F., & Schuurman, N. K. (2019). The challenge of generating causal hypotheses using network models. https://doi.org/10.31234/osf.io/ryg69

Safavian, N. (2019). What makes them persist? Expectancy-value beliefs and the math participation, performance, and preparedness of hispanic youth. *AERA Open, 5*, 233285841986934. https://doi.org/10.1177/2332858419869342

Safavian, N., & Conley, A. (2016). Expectancy-value beliefs of early-adolescent hispanic and non-hispanic youth: Predictors of mathematics achievement and enrollment. *AERA Open, 2*, 233285841667335. https://doi.org/10.1177/2332858416673357

Sanson-Fisher, R. W., Bonevski, B., Green, L. W., & D'Este, C. (2007). Limitations of the randomized controlled trial in evaluating population-based health interventions. *American Journal of Preventive Medicine, 33*(2), 155–161. https://doi.org/10.1016/j.amepre.2007.04.007

Schloss, P. D. (2018). Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *MBio, 9*(3), e00525–e00518.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality, 47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–278). The Guilford Press.

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: a longitudinal examination of the links between choices and beliefs. *Developmental Psychology, 42*(1), 70–83. https://doi.org/10.1037/0012-1649.42.1.70

Simpkins, S. D., Fredricks, J. A., & Eccles, J. S. (2012). Charting the Eccles' expectancy-value model from mothers' beliefs in childhood to youths' activities in adolescence. *Developmental Psychology, 48*, 1019–1032. https://doi.org/10.1037/a0027468

Spinath, B., & Steinmayr, R. (2008). Longitudinal analysis of intrinsic motivation and competence beliefs: Is there a relation over time? *Child Development, 79*(5), 1555–1569. https://doi.org/10.1111/j.1467-8624.2008.01205.x

Steinmayr, R., & Spinath, B. (2008). Sex differences in school achievement: what are the roles of personality and achievement motivation? *European Journal of Personality, 22*(3), 185–209. https://doi.org/10.1002/per.676

Stevenson, H., & Stigler, J. W. (1994). *Learning gap: Why our schools are failing and what we can learn from Japanese and Chinese educ*. Simon and Schuster.

Tang, X., Lee, H. R., Wan, S., Gaspard, H., & Salmela-Aro, K. (2022). Situating expectancies and subjective task values across grade levels, domains, and countries: A network approach. *AERA Open, 8*, 23328584221117168. https://doi.org/10.1177/23328584221117168

Trautwein, U., Marsh, H. W., Nagengast, B., Lüdtke, O., Nagy, G., & Jonkmann, K. (2012). Probing for the multiplicative term in modern expectancy–value theory: a latent interaction modeling study. *Journal of Educational Psychology, 104*(3), 763–777. https://doi.org/10.1037/a0027470

Usher, E. L. (2018). Acknowledging the whiteness of motivation research: seeking cultural relevance. *Educational Psychologist, 53*(2), 131–144. https://doi.org/10.1080/00461520.2018.1442220

Viljaranta, J., Tolvanen, A., Aunola, K., & Nurmi, J. E. (2014). The developmental dynamics between interest, self-concept of ability, and academic performance. *Scandinavian Journal of Educational Research, 58*(6), 734–756. https://doi.org/10.1080/00313831.2014.904419

Waldorp, L. (2020). Network discovery and causal analysis. https://osf.io/mpwc7/. Accessed 12/10/20.

Wang, M. T., & Eccles, J. S. (2013). School context, achievement motivation, and academic engagement: a longitudinal study of school engagement using a multidimensional perspective. *Learning and Instruction, 28*, 12–23.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "$p < 0.05$." *The American Statistician, 73*(sup1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Weidinger, A. F., Steinmayr, R., & Spinath, B. (2017). Math grades and intrinsic motivation in elementary school: a longitudinal investigation of their association. *British Journal of Educational Psychology, 87*(2), 187–204. https://doi.org/10.1111/bjep.12143

Weidinger, A. F., Spinath, B., & Steinmayr, R. (2019). The value of valuing math: longitudinal links between students' intrinsic, attainment, and utility values and grades in math. *Motivation Science, 6*(4), 413–422. https://doi.org/10.1037/mot0000179

Whitaker, K. (2017). Publishing a reproducible paper. https://doi.org/10.6084/m9.figshare.5440621.v2

Wigfield, A. (1994). Expectancy—value theory of achievement motivation: a developmental perspective. *Educational Psychology Review, 6*, 49–78.

Wigfield, A., & Eccles, J. S. (2020). 35 years of research on students' subjective task values and motivation: A look back and a look forward. In *Advances in motivation science* (Vol. 7, pp. 161–198). Elsevier.

Wigfield, A., Tonks, S., & Klauda, S. L. (2009). Expectancy-value theory. *Handbook of motivation at school* (pp. 69–90). Routledge.

Wigfield, A., Muenks, K., & Eccles, J. S. (2021). Achievement motivation: what we know and where we are going. *Annual Review of Developmental Psychology, 3*, 87–111. https://doi.org/10.1146/annurev-devpsych-050720-103500

Xu, J. (2018). Reciprocal effects of homework self-concept, interest, effort, and math achievement. *Contemporary Educational Psychology, 55*, 42–52. https://doi.org/10.1016/j.cedpsych.2018.09.002

Yoon, K. S. (1996). *Testing reciprocal causal relations among expectancy, value and academic achievement of early adolescents: A longitudinal study* (Order No. 9635640) (Publication No. 9635640) [Doctoral Dissertation, The University of Michigan]. ProQuest Dissertations & Theses Global.

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020). From data to causes I: building a general cross-lagged panel model (GCLM). *Organizational Research Methods, 23*(4), 651–687. https://doi.org/10.1177/1094428119847278