



# Characterizing the psychosocial effects of participating in a year-long residential research-oriented learning community

Alejandra J. Magana<sup>1</sup> · Aparajita Jaiswal<sup>1</sup> · Aasakiran Madamanchi<sup>2</sup> · Loran C. Parker<sup>3</sup> · Ellen Gundlach<sup>4</sup> · Mark D. Ward<sup>5</sup>

Accepted: 10 March 2021 / Published online: 23 March 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Research on learning communities has primarily focused on identifying institutional outcomes such as student achievement and retention. However, more research is needed on how the learning community experience impacts the motivation, beliefs, and perceptions associated with student success. This study investigates the psychosocial effects of participating in a residential research-oriented learning community regarding students' interest and motivation in pursuing research-oriented careers, research and data self-efficacy beliefs, sense of belongingness with the learning community, and socialization levels and career awareness in research-oriented fields. This study also investigated the mediating effects of students' initial research self-efficacy beliefs on differential gains regarding career awareness, motivation and interest, and sense of belongingness and socialization after one year of participating in a residential research-oriented learning community. Participants of the study consisted of five cohorts of the learning community, each composed of twenty students. Students in each cohort participated in a pretest-posttest design survey study. Findings suggest that alignment of student interest with the learning community discipline is a key mediator of student growth in their self-efficacy beliefs, sense of belongingness with the learning community and levels of socialization, and career awareness in the selected field. Implications include recommendations for the thoughtful design of learning communities that promote cognitive apprenticeships by orchestrating the content, method, sequencing, and sociology of the learning environment.

**Keywords** Learning communities · Self-efficacy beliefs · Motivation and interest · Undergraduate research · Career awareness

## Introduction

Learning communities create a confluence of scholastic and social interaction as they bring together students sharing the same interest under one umbrella. Learning communities often bring together groups of college students in cohorts who take linked courses together (Price, 2005). Colleges and universities are expanding learning communities' implementation to forge closer

bonds between students, students and faculty, and students and the institution (Price, 2005). The ultimate goal is that the students' learning experience is enhanced when "the learning that evolved from these communities is collaborative, in which the collaborative knowledge of the community is greater than any individual knowledge" (Johnson, 2001, p. 34). Learning communities can have different configurations such as curricular learning communities, classroom learning communities,

✉ Alejandra J. Magana  
admagana@purdue.edu

Aparajita Jaiswal  
jaiswal2@purdue.edu

Aasakiran Madamanchi  
amadaman@purdue.edu

Loran C. Parker  
carleton@purdue.edu

Ellen Gundlach  
gundlach@purdue.edu

Mark D. Ward  
mdw@purdue.edu

<sup>1</sup> Computer and Information Technology, Purdue University, West Lafayette, IN 47906, USA

<sup>2</sup> Polytechnic Institute, Purdue University, West Lafayette, IN 47906, USA

<sup>3</sup> Evaluation and Learning Research Center, Purdue University, West Lafayette, IN 47906, USA

<sup>4</sup> The Data Mine, Purdue University, West Lafayette, IN 47906, USA

<sup>5</sup> Statistics and The Data Mine, Purdue University, West Lafayette, IN 47906, USA

residential learning communities, and student-type learning communities (Lenning & Ebberts, 1999). However, what all of them have in common is that all are forms of communities of practice. Communities of practice (Wenger, 2011; Wenger & Snyder, 2000) are groups of people who engage in collective learning in a shared domain of interest.

Research has revealed that learning communities are an effective learning initiative that can improve student retention, surges in academic achievement, diminished faculty isolation, and increased curricular integration (Lenning & Ebberts, 1999). However, while universities are promoting and investing in the concept of community to enhance the student learning experience, there is still a need for more evidence on their effectiveness, focusing on the student experience (Dawson et al., 2006). The majority of research has focused on institutional outcomes, but little has focused on how the learning community experience impacts the motivation, beliefs, and perceptions associated with student success (Barefoot, 2000; Browne & Minnick, 2005). To contribute to the body of research focused on students' experiences in learning communities, this study identifies the effectiveness of a learning community that uniquely integrates four academic and socialization elements: (1) a learning community comprising of shared coursework in the domain of data science; (2) a faculty-mentored, team-oriented data science research experience; (3) a living community in the same residence hall with social and academic activities, and (4) participation in a professional development seminar for career exploration in data science. This study identifies the psychosocial effects of participating in this learning community in terms of motivation and interest, research self-efficacy, sense of belongingness and socialization in this context, and career awareness in the field. The study's specific aims are to first identify the psychosocial effects of participating in a residential learning community regarding students' interest and motivation on pursuing research-oriented careers, research self-efficacy beliefs, sense of belongingness with the learning community, and socialization, and career awareness in research-oriented fields. Second, to describe how student levels of research self-efficacy beliefs correspond to differential gains on career awareness, motivation and interest, and sense of belongingness and socialization in the learning community.

## Learning Communities

Constructivist learning theories, including situated learning theory and communities of practice (Bielaczyc & Collins, 1999; Brown & Campione, 1994; Daly et al., 2003), define learning communities as curated academic and social learning experiences designed to support peer-to-peer and student-faculty interaction (Jessup-Anger, 2015). Learning communities, which have become a mainstay of the higher education

landscape over the last three decades, exist in various forms (Fink & Inkelas, 2015). The earliest learning communities focused on providing bridge experiences to acclimate first-year students to college (Garcia, 1991; Tinto & Goodsell, 1994). Many learning communities in this tradition are focused on supporting students' retention and success from groups at risk of leaving college (Contreras, 2011; Tinto & Engstrom, 2008). Some studies have demonstrated greater retention, academic performance, and self-reported engagement from first-year students in learning communities (Baker & Pomerantz, 2000; Taylor et al., 2003; Tokuno & Campbell, 1992; Zhao & Kuh, 2004).

These positive effects have largely been attributed to enhanced social relationships among students (Carrino & Gerace, 2016; Franklin, 2000; MacGregor, 1991; Tinto & Russo, 1994; Zhao & Kuh, 2004). Studies have documented that learning communities foster increased student-to-student and student-faculty interactions (Baker & Pomerantz, 2000; Kuh, 2008; Pike, 1999). Indeed, these increased social interactions appear to generate a sense of belonging among students; learning community participants report greater feelings of affiliation and inclusivity (Dodd, 2002). Also, living or residential learning communities intentionally focus on combining students' residence hall experience with curricular and co-curricular experiences. The goal is to build strong connections between the academic and social aspects of college life and at the same time enable formal and informal forms of peer learning (Inkelas et al., 2008).

Despite the abundance of research on the effects of learning communities, more work is needed that investigates the effectiveness of specific learning community initiatives. For example, most learning community research has focused on large first-year experience-based learning communities. But there has been limited study of increasingly popular discipline-specific learning communities (Dagley et al., 2016; Solanki et al., 2019). Dagley et al. have reported that a residential learning community for STEM students led to greater first-year retention, long-term retention, and graduation than a comparison group (Dagley et al., 2016). Solanki and colleagues showed increased academic performance (first-year GPA) among participants in a bioscience-focused learning community (Solanki et al., 2019). Additionally, studies of learning communities have remained focused on quantifiable aggregate statistics such as GPA, retention, and graduation rates (Holt & Nielson, 2019; Lardner & Malnarich, 2009). A more granular study characterizing differential outcomes of students within learning communities is mostly lacking.

The implications of the previous work on learning communities for this study's design relate to measuring the psychosocial effects of participating in a living-learning community. The primary goal is to engage students in collaborative activities aimed at developing research and data science skills and increasing their awareness of career occupations in this field.

## Psychosocial Effects of Participation in Learning Communities

Participating in learning communities provides students with learning opportunities to develop their knowledge and skills to solve real-world problems (Solanki et al., 2019). However, it is equally important for learners to build confidence in mastering a concept or skill, be self-motivated to learn, and apply the acquired knowledge in other contexts (Peters-Burton et al., 2015). Such skills can be generally characterized as forms of psychosocial factors or effects, herein defined when one's views or beliefs about the self (psychologically or socially) are being affected (Carrino & Gerace, 2016). Specifically, we focus on psychosocial effects such as self-belief, motivation and interest, sense of belonging, and career awareness. For this, we grounded our study on Bandura's theory of self-efficacy (1994). Students' self-efficacy beliefs, their interest, and motivation are highly related (Schunk et al., 2012). Self-efficacy is defined as "people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives" (Bandura, 1994, p. 71). Motivation refers to the activation to action; "the level of motivation is reflected in the choice of courses of action, and the intensity and persistence of effort." (Bandura, 1994, p. 71). Self-efficacy thus determines how individuals motivate themselves and consequently behave (Bandura, 1994). Self-efficacy beliefs include one's ability to plan for and execute the steps necessary for future success (Bandura, 1977).

When students believe in their efficacy to achieve a task, they become motivated to perform in ways that make their achievement more likely (Bandura, 1977, 2006). Similarly, self-efficacy and sense of belonging are also known predictors of motivation and performance (Blaney & Stout, 2017; Walton & Cohen, 2007). Sense of belonging refers to a subjective personal sense of fitting in and being included as a valued and legitimate member of an academic discipline (Goodenow, 1993). Furthermore, students' beliefs about their capabilities in disciplinary areas have also been shown to influence their career choice (Zeldin et al., 2008). These relationships can be explained under Bandura's views that people's motivation and actions are influenced more by what they believe than what is objectively true (Bandura, 1994).

Research that investigated how self-efficacy and other psychosocial variables might jointly be affected by participating in learning communities include the work from Freeman et al. (2008). They identified that participating in linked classes as part of a learning community positively influenced students' attitudes, learning experiences, and intrinsic motivation in STEM (Freeman et al., 2008). Similarly, in the context of a virtual learning community (Sun et al., 2012), it was hypothesized that task complexity and self-efficacy – two social learning factors – moderate the relationship between motivation and sustained participation. As part of their findings,

researchers identified that motivation significantly influenced sustained participation intention (Sun et al., 2012).

Implications from previous work on living-learning communities (Parker & Ward, 2019) and previous work on the effects of different socialization forms in learning higher education suggested the need for further research. This study contributes to research on disciplinary learning communities by investigating two research questions. (1) What are the psychosocial effects of participating in a residential learning community, such as on students' interest and motivation in pursuing research-oriented careers, research and data self-efficacy beliefs, sense of belongingness and socialization within the learning community, and career awareness in research-oriented fields? And (2) How do student entering profiles regarding research self-efficacy beliefs correspond to differential gains on career awareness, motivation and interest, and sense of belongingness and socialization in the learning community?

## Methods

This study used a pretest-posttest design survey study to the effects of participating in a living-learning community consisting of a year-long, cohort-based program to introduce current statistics majors and interested students from other majors to data science practices, educational and career pathways in statistics and data science, and applied research utilizing statistics and data science.

## Context and Participants

The study's context focuses on a statistics and data science learning community at a large midwestern university in the United States of America, consisting of a year-long, cohort-based program. This program was conceived as a bridge program between the strong existing first-year experience at the university and the more variable second-year experiences as students move into more specialized courses in their disciplines. The program's holistic approach includes (1) a learning community (block-scheduling) comprising three courses on computing with data, probability, and statistics; (2) a faculty-mentored, team-oriented, data science research experience involving applied investigations appropriate for sophomores; (3) a living community in the same residence hall floor, with social and academic activities to enhance the sophomore experience; and (4) participation in a professional development seminar each semester, to prepare for graduate school and career exploration in data science.

Participants of the study consisted of five cohorts of the learning community, each composed of twenty students. For each year of the program, sophomore students about to start their second year in the statistics major were invited to apply

to be part of the learning community. About 55 students applied each year, and from those, 20 were selected to participate. There were no specific selection criteria, but for each candidate, administrators followed a holistic review process aimed at identifying those students for whom this would be an impactful and meaningful experience. Students in each cohort participated in a voluntary assessment at the start and the end of the semester. For this study, we considered all the 100 students who participated in the learning community between the Fall of 2014 until the Fall of 2018. From the total of 100 students, 84 of them responded to all the questions in the survey; therefore, we considered 84 students as the sample of our study.

### Data Collection Method and Procedures

The data collection method consisted of a 41-question survey that was administered to all participating students at the beginning and at the end of the academic year herein called the Statistics and Data Science Learning Community Student Survey (SDSLCSS). The Likert-scale survey covered four main constructs: Interest and motivation, research self-efficacy, sense of belongingness and socialization, and career awareness (Appendix Table 4). Two investigators developed this survey, an expert in statistical research and an expert in education research and evaluation. The survey instrument was created by adopting and adapting survey items from literature in motivation, self-efficacy, undergraduate research, and career awareness. As the sample size for the study was too small to allow for construct validity using factor analysis (Anthoine et al., 2014; Williams et al., 2010), we used a content validity approach, described in the following section. Content validation is widely used in education and is particularly appropriate for the validation of new instruments (MacKenzie et al., 2011; Podsakoff et al., 2016). Further, instrument reliability was assessed by measuring the internal consistency of each construct, as described in the following section.

### Content Validity Analysis & Internal Reliability for the SDSLCS Scale

Content validation is an effective technique that requires an expert's judgment to validate the constructs or themes that emerge from an assessment or test (Crocker, 2001). Experts' knowledge and proficiency can determine the construct relevance, representation, and quality of the instrument to measure a particular domain or dimension (García-Valderrama & Mulero-Mendigorry, 2005; Messick, 1995). The content validation method was conducted in a two-step process. Three experts, two education researchers, and one statistics researcher were provided with the four-domains, eleven subdomain definitions, and items represented under each domain. The experts were asked to critically review each domain and

subdomain definitions and all items under each domain, and independently score the items on a four-point relevance scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, 4 = highly relevant) (Polit & Beck, 2006). Then two types of content validity scores were calculated; the *content validity Index score for items* (I-CVI) and the *content validity score for the scale* (S-CVI). The I-CVI is the proportion of experts who rate any given item as a 3 or 4 on a 4-point scale (DeVon et al., 2007, p. 158). The I-CVI scoring revealed that 41 items were rated as either 3 or 4 on the relevance scale by all three raters. The S-CVI, the proportion of valid items in the scale, is 0.86 (DeVon et al., 2007, p. 158). The CVI scores obtained for items (I-CVI = 1) and scales (S-CVI = 0.86) meet the acceptable criteria for content validity in the case of three raters (Polit & Beck, 2006). Table 1 presents the domain definition, subdomains, and the number of items. Also, refer to Appendix Table 4 to see the complete survey used for the study.

We calculated the internal consistency (Cronbach's alpha) for the items within each of the four constructs measured with the survey instrument to account for reliability. Table 2 presents the results for internal consistency. According to the scores, reliability was considered acceptable. The generally accepted value for Cronbach's alpha ( $\alpha$ ) denotes that  $\alpha = 0.6$  to  $0.7$  implies an acceptable level of reliability,  $\alpha = 0.8$  or greater determines a very good reliability level (Ursachi et al., 2015). Based on the acceptable level of alpha values, we concluded that the instrument as a whole ( $\alpha = 0.88$ ) and the three themes of motivation/interest, self-efficacy, and career awareness (see. Table 2) demonstrated a very good level of internal consistency. For the case of the sense of belonging, alpha values demonstrated an acceptable level of internal consistency.

### Data Analysis

Descriptive and inferential analyses were used to describe the students' psychosocial effects of a year-long learning community experience. Scores were rescaled between 0 and 1, and those were interpreted as follows: scores between 0 and 0.33 were interpreted as low or negative levels of psychosocial effects, scores between 0.34 and 0.66 were considered as moderate psychosocial effects, and between 0.67 and 1 were considered as high or positive levels of psychosocial effects.

Additionally, we used a clustering algorithm to identify groups of students that demonstrate high intra-class homogeneity and high inter-class heterogeneity (Battaglia et al., 2015). Clustering is an unsupervised machine learning technique, which is particularly useful for grouping unlabeled/unclassified data (Kogan, 2007). Prior studies have demonstrated the use of a hierarchical clustering algorithm to group the students based on their responses on a multiple-choice test (Ding & Beichner, 2009) and survey instruments (Medová &

**Table 1** Psychosocial effects of learning community participation: Domains and Subdomains

Domain	Domain Definition	Subdomains	Sample Questions
Motivation/ Interest (M/I)	Desire to become a researcher and intentions to choose a research-oriented career	Motivation, n=(2)	Q1: I possess the motivation and persistence required for a career in a research-oriented field
		Interest, n=(4)	Q3: I have a strong interest in pursuing a career as a researcher
Self-Efficacy (SE)	Perceived competence or ability to engage in research-related activities such as writing skills, data analysis skills, research planning skills, academic competence, problem solving competence, engage in research with others, and research knowledge.	SE: Pursue a research career, n=(1)	Q7: I have the ability to pursue a successful career as a researcher
		SE: Research knowledge, n=(9)	Q11: I have the ability to understand the research process
		SE: Research planning skills, n=(3)	Q17: I have the ability to organize my research ideas in writing
		SE: Data analysis and interpretation skills, n=(6)	Q22: I have the ability to statistically analyze data using computer software
		SE: Data communication skills, n=(2)	Q26: I have the ability to orally communicate the results of research projects
		SE: Research application skills, n=(1)	Q28: I have the ability to apply what I learn in class to real world problems
		SE: Groupwork skills, n=(4)	Q29: I can work collaboratively with others
Career Awareness (CA)	Knowledge and interest in research-oriented career opportunities.	CA: Career awareness, n=(3)	Q34: I am aware of the various research-oriented career opportunities available to me
Sense of Belonging (SOB)	Sense of belongingness to an institution, sense of belongingness to a class, perceived institutional support and access to resources, and comfort level of socialization	SOB: Institution, n=(1)	Q36: I feel a sense of belonging at [Blinded]
		SOB: Institutional support, n=(1)	Q 39: I have social support at [Blinded]
		SOB: with others, n=(2)	Q41 I participate in study groups

Bakusová, 2019), for identifying significant and useful clusters. For our study, we specifically used Ward's minimum variance clustering (Ward Jr, 1963), a type of hierarchical clustering algorithm to cluster students' self-efficacy domain responses on the SDSLCSS for identifying significant student groups from a limited sample size ( $N = 84$ ). A prior study by (Antonenko et al., 2012) demonstrated the use of Ward's minimum variance algorithm to cluster a small sample of 59 students enrolled in an introductory instructional technology course. Specifically, Antonenko et al., (2012) used hierarchical clustering to divide the students into four groups of unequal sizes: Cluster 1 ( $n = 13$ ), Cluster 2 ( $n = 7$ ), Cluster 3 ( $n = 19$ ), and Cluster 4 ( $n = 20$ ). Further, they conducted one-way ANOVA for the clusters as between-subject factors and considered problem-solving performance as the dependent variable. The results from Antonenko et al., (2012) revealed that there was a significant difference among the clusters.

Hierarchical clustering is a commonly used method for the small sample size. For example, a study by (Medová & Bakusová, 2019) used Ward's minimum variance clustering to group 30 in-service mathematics teachers based on their response to a questionnaire related to teacher's beliefs and

current pedagogy. The studies also revealed that a hierarchical clustering method using Ward's minimum variance is an appropriate clustering method when there is no preconceived notion about the number of clusters that could be formed from a particular dataset. Since the analysis was exploratory and the data set was small, we found Ward's minimum variance clustering algorithm appropriate for our study. The student responses to the questions related to the self-efficacy domain for T1 and T2 were then clustered using Ward's minimum variance method. Since Ward's minimum variance method follows an agglomerative strategy of hierarchical clustering, at every step of Ward's minimum variance clustering, a metric

**Table 2** Psychosocial participation effects domain reliability

Themes	No. of Items	Alpha Value
Motivation/Interest	6	0.84
Self-Efficacy	26	0.86
Career Awareness	3	0.84
Sense of Belonging	6	0.72

is computed, which is the sum of Euclidean distance of each student response from its cluster's mean. Further, the algorithm combines the different clusters sequentially to find the pairs of clusters to be merged, to minimize the increase in the sum of Euclidean distance of each student response from their cluster's mean. The output of Ward's minimum variance approach resulted in three clusters:

Cluster 1: is the group of students that demonstrated a *high* self-efficacy, and *high* motivation and interest

Cluster 2: is the group of students that demonstrated *moderate* self-efficacy, and *moderate* motivation and interest

Cluster 3: is the group of students that demonstrated *moderate* self-efficacy, and *high* motivation and interest

To verify the number of clusters, we used the elbow method (Yuan & Yang, 2019). The elbow method aims to identify the optimal number of clusters sufficient to explain the variance in the observations; adding extra clusters does not significantly improve the model's ability to explain variability in data. In our case, using the elbow curve, we found that three clusters were optimal (see Appendix 2). Further, based on a Welch's Test ( $F(2, 47) = 42.46, p < .001$ ), it was identified that the clusters were statistically significantly different within each domain and subdomain over time. Finally, survey question responses were rescaled to allow valid comparison across questions.

## Results

The results are organized into two main sections. The first section addresses the first research question by providing an overview of trends regarding students' psychosocial effects in a statistical learning community. The second section describes three distinct clusters of students based on their initial research self-efficacy beliefs who participated in the learning community and the different growth patterns of these student clusters during their participation in the learning community. Comparison of these clusters in the second section addresses the second research question regarding the role of research self-efficacy beliefs and their relationship to career awareness, motivation and interest, and sense of belonging and socialization in the statistics learning community.

### Psychosocial Effects of Student Participation in a Statistical Learning Community

As shown in Table 3, at the beginning of the year (T1), the learning community students ranged between moderate and high in all psychosocial domain scores. Scores in the high range included research self-efficacy, motivation and interest,

and a sense of belonging. Only the Career Awareness domain had an average student score in the moderate range. By the end of the year (T2), the average student score for participating students indicated high or positive knowledge and self-beliefs in career awareness, research-self efficacy, and sense of belonging and socialization. A paired t-test was used to compare domain scores from the two time-points. The analysis revealed that the mean scores of the data science learning community students' scores increased significantly in all domains other than the motivation and interest in research fields (Table 3). This domain stayed relatively constant over the year.

### Cluster Analysis of Students

Cluster analysis was used to uncover students' subgroups with differential responses to participation in the learning community. Specifically, a cluster analysis on research self-efficacy scores uncovered three distinctive groups (see Fig. 1).

One group (cluster 1,  $n = 41$ ) started with high research self-efficacy, high motivation and interest, and a high sense of belonging. The second group (cluster 2,  $n = 26$ ) started with moderate research self-efficacy, moderate motivation and interest, and a high sense of belonging. The third group (cluster 3,  $n = 17$ ) started with moderate self-efficacy, high motivation, and a high sense of belonging. Each of the three groups is described in the following sections as well as their post-participation psychosocial effects.

Cluster 1: *High* self-efficacy and *high* motivation and interest

Cluster 1, as observed in Fig. 2, is characterized by those students who, at the time they entered the learning community, primarily exhibited high levels of research self-efficacy ( $M = .75, SD = .06$ ), high levels of motivation and interest ( $M = .74, SD = .13$ ), and high levels of sense of belonging and socialization ( $M = .83, SD = .11$ ). These students also exhibited a moderate level of career awareness in research ( $M = .56, SD = .19$ ). These students, after one year of having participated in the learning community, reported high increases on each of the psychosocial domains including research self-efficacy ( $M = .85, SD = .06, t(40) = 7.32, p < .001$ ), motivation and interest ( $M = .75, SD = .15, t(40) = 0.43, p = .67$ ), career awareness in research-related fields ( $M = .87, SD = .12, t(40) = 10.16, p < .001$ ), and sense of belonging and socialization ( $M = .89, SD = .10, t(40) = 3.69, p < .001$ ). From these four domains, the increases were statistically significant except for the measure of motivation and interest, which remained constant.

Appendix Table 5 shows the subdomain scores for the students in cluster 1 at both time-points and the change and statistical significance. The largest changes in research self-

**Table 3** Aggregate measures of psychosocial domains of student participation in a statistical learning community

	T1		T2		Gain	df	t	P-Val
	Mean	SD	Mean	SD				
Self-efficacy (SE)	0.69	0.08	0.81	0.09	0.12	83	11.27	0.000
Motivation and interest (MI)	0.71	0.14	0.73	0.17	0.01	83	0.89	0.38
Career awareness (CA)	0.55	0.17	0.82	0.14	0.27	83	12.68	0.000
Sense of belonging (SOB)	0.80	0.11	0.86	0.11	0.07	83	5.59	0.000

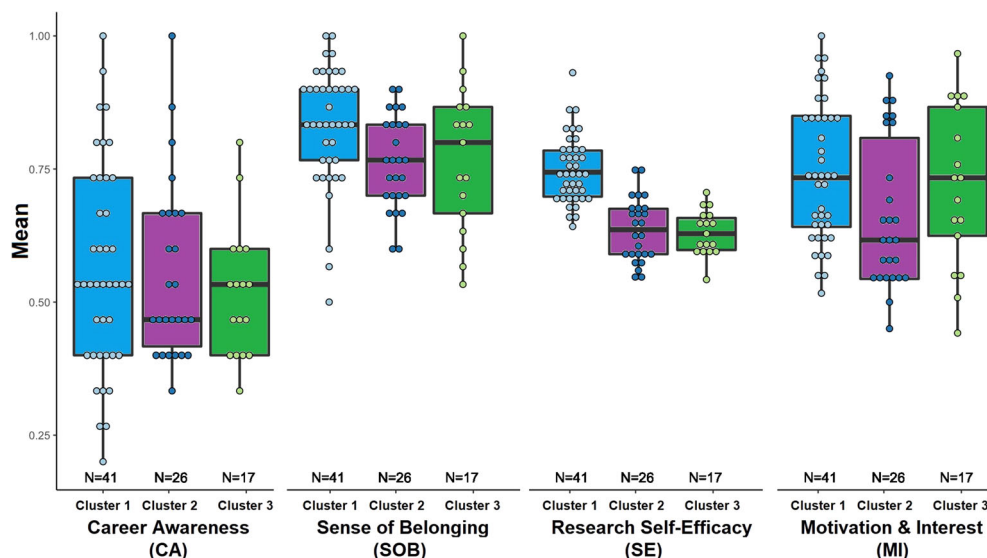
efficacy were related primarily to data analysis and interpretation skills and data communication skills, and overall research knowledge. Research planning skills, group work skills, and the ability to pursue a research-oriented career changed only minimally. Students in Cluster 1 also experienced large changes in career awareness. Regarding their sense of belonging and socialization, students experienced the largest gain in terms of a sense of belonging with their peers. Finally, it can be observed that this group of students started with high levels of motivation and interest, which remained unchanged throughout their participation in the year-long experience.

*Cluster 2: Moderate self-efficacy and moderate motivation and interest*

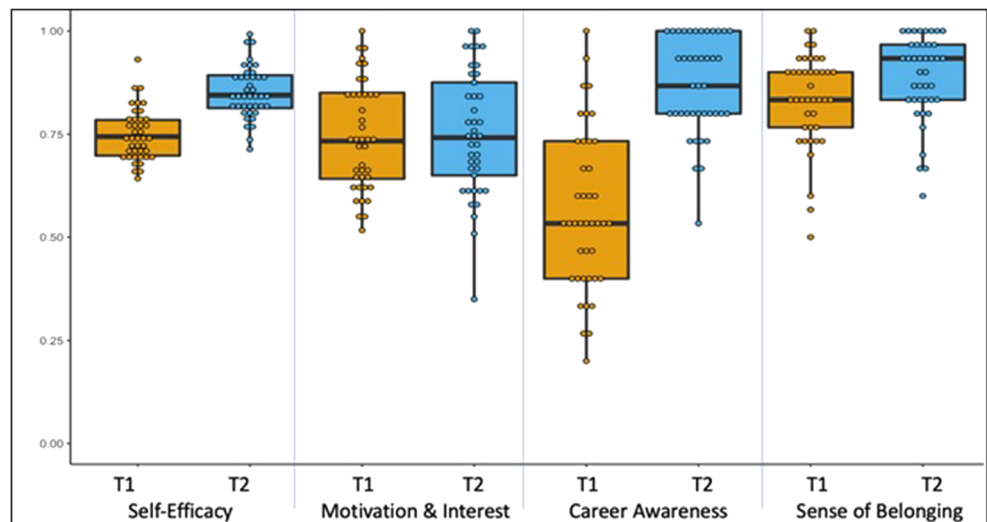
Cluster 2, as observed in Fig. 3, is characterized by those students who, at the time they entered the learning community, primarily exhibited moderate levels of research self-efficacy ( $M = .64$ ,  $SD = .06$ ), moderate levels of motivation and interest ( $M = .66$ ,  $SD = .14$ ), and high levels of sense of belonging and socialization ( $M = .76$ ,  $SD = .09$ ). These students also exhibited a moderate level of career awareness in

research ( $M = .55$ ,  $SD = .17$ ). After one year of having participated in the learning community these students reported higher levels in the psychosocial domains of research self-efficacy ( $M = .71$ ,  $SD = .04$ ),  $t(25) = 5.87$ ,  $p < .001$ , and career awareness in research-related fields ( $M = .71$ ,  $SD = .15$ ),  $t(25) = 4.67$ ,  $p < .001$ . This group also demonstrated higher levels of sense of belonging and socialization ( $M = .80$ ,  $SD = .09$ ),  $t(25) = 1.91$ ,  $p = .07$ , and lower levels of motivation and interest ( $M = .64$ ,  $SD = .16$ ),  $t(25) = -1.01$ ,  $p = .32$ , but these changes did not reach statistical significance.

From Appendix Table 6, it can be observed that the largest changes in research self-efficacy were related primarily to the subdomains of research knowledge, data analysis/interpretation skills, and research communication. However, this group also experienced a statistically significant decrease in their perceived ability to pursue a research-oriented career. In general, students also experienced large changes in career awareness regarding research-oriented career opportunities, research in graduate school, and other career options they could specialize in. Regarding their sense of belonging and socialization, students experienced equal gains in terms of perceived institutional support, socialization, and sense of belonging with their peers, but not changes in the institution's

**Fig. 1** Psychosocial effects on domain scores at T1 for each cluster

**Fig. 2** Psychosocial effects experienced by students in Cluster 1 at T1 and T2



perceived sense of belonging. Finally, it can be observed that this group of students started with moderate levels of motivation and interest, which slightly decreased after the year-long experience but not significantly.

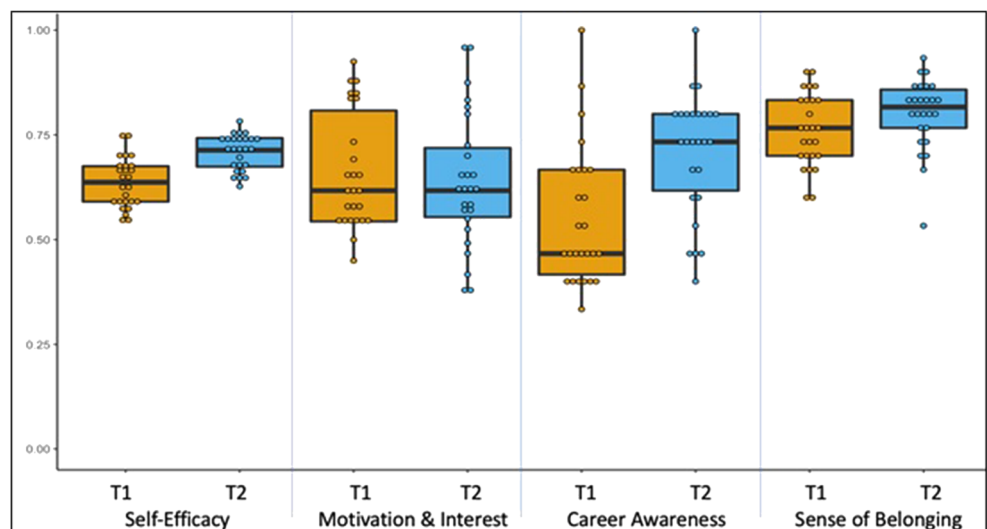
**Cluster 3: Moderate self-efficacy and high motivation and interest**

Cluster 3, as observed in Fig. 4, can be characterized by those students who, at the time they entered the learning community, primarily exhibited moderate levels of research self-efficacy ( $M = 0.63, SD = 0.04$ ), high levels of motivation and interest ( $M = .72, SD = .15$ ), and high levels of sense of belonging and socialization ( $M = .76, SD = .09$ ). These students also exhibited a moderate level of career awareness in research ( $M = .77, SD = .13$ ). These students, after one year of having participated in the learning community, reported higher levels on each of the psychosocial domains including research self-

efficacy ( $M = .85, SD = .06, t(16) = 11.42, p < .001$ ), motivation and interest ( $M = .79, SD = .18, t(16) = 2.60, p = .02$ ), career awareness in research-related fields ( $M = .87, SD = .10, t(16) = 9.05, p < .001$ ), and sense of belonging and socialization ( $M = .90, SD = .11, t(16) = 4.52, p < .001$ ). This is the only group of students who on average, experienced statistically significant increases in the four psychosocial domains.

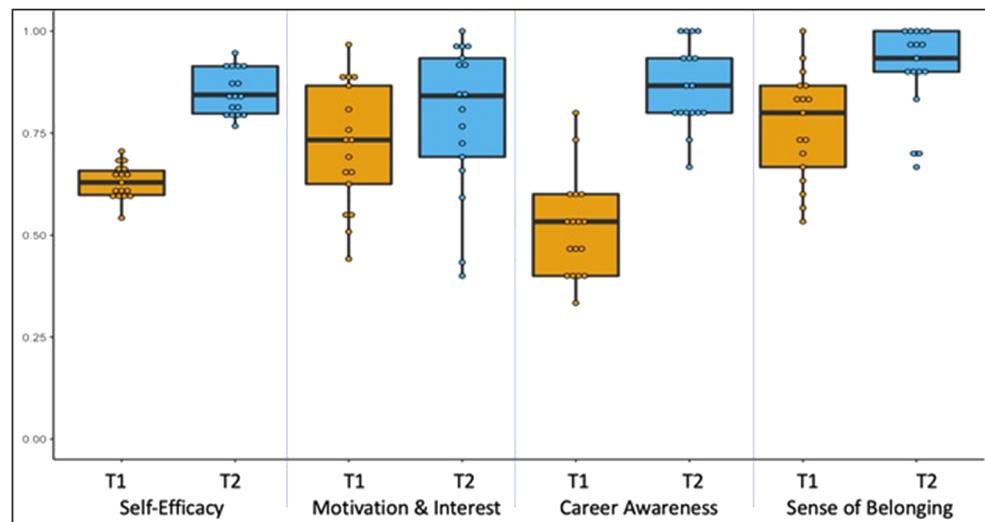
From Appendix Table 7, it can be observed that the largest changes in research self-efficacy were related primarily to the subdomains of research knowledge, followed by research communication skills. Then, gains were about the same for research planning skills, research application skills to solve real problems, and data analysis and interpretation skills. In general, students also experienced large changes in career awareness regarding research-oriented career opportunities, research in graduate school, and other career options they could specialize in. Finally, regarding their sense of belonging and socialization, students experienced the largest gains in

**Fig. 3** Psychosocial effects experienced by students in Cluster 2 at T1 and T2





**Fig. 4** Psychosocial effects experienced by students in Cluster 3 at T1 and T2



their perceptions of socialization and a sense of belonging with their peers, but no statistical gains regarding the perceived institutional support. Finally, it can be observed that although this group of students started with high levels of motivation and interest, they still statistically increased their perceptions in this regard.

## Discussion and Implications

Findings from our study aimed to understand factors that contribute to student positive experiences in the context of a year-long, research-oriented living-learning community. In this context, positive experiences can be characterized by students' overall high levels of interest and motivation, research self-efficacy beliefs, sense of belongingness and socialization, and moderate levels of career awareness of related fields. Overall, we identified that a learning community that integrates (1) block-scheduling comprising three courses on computing with data, statistics, probability; (2) a faculty-mentored, team-oriented, data science research experience involving applied investigations appropriate for sophomores; (3) a living community in the same residence hall floor, with social and academic activities to enhance the sophomore experience; and (4) participation in a professional development seminar each semester, to prepare for graduate school and career exploration in data science, can result in positive experiences for the students. This positive trend can be observed from overall increases in the average scores regarding the whole sample's four psychosocial effects (see Appendix 6).

However, we also identified subgroups of students that on average, entered into the learning community with different levels of research self-efficacy beliefs, interest and motivation, and sense of belonging and socialization. For instance, regarding research self-efficacy beliefs, one group of students

(Cluster 1) entered the learning community with high levels of self-efficacy, with statistically significant increases at the end of the year-long experience. The other two groups of students (Cluster 2 and Cluster 3) entered the learning community with moderate self-efficacy levels and experienced statistically significant increases in their research self-efficacy at the end of the year-long experience. However, Cluster 3 that started at the same self-efficacy level as Cluster 2, reached parity with the high self-efficacy cluster 3 by the end of the year. In contrast, Cluster 2, despite its statistically significant increase, failed to reach even the starting level of Cluster 1. Additionally, the difference in the self-efficacy endpoints of Cluster 1 and 3 versus Cluster 2 may be attributed to a 'ceiling effect' in the self-efficacy questions.

A similar pattern played out in the sense of belongingness and socialization domain. Although all three clusters demonstrated statistically significant growth over the year, they did not have congruent responses. As in the self-efficacy domain, students in Clusters 2 and 3 entered the year at lower levels of sense of belonging than students in Cluster 1. By the end of the year, Clusters 1 and 3 had reached 'near saturation' levels of sense of belonging. Similar to the self-efficacy domain, students in Cluster 2, despite its improvement, did not reach the starting point of students in Cluster 1. Interestingly, part of these effects seems to be mediated by a change in the subdomains of sense of belonging. For this subdomain, Cluster 3 started in the moderate range with Cluster 2, but ended with a higher mean than Cluster 1, whereas Cluster 2 had a relatively modest improvement. This pattern continued within the sense of belonging to the institution subdomain, but in this category, all three clusters were relatively high, so the differences were muted. The three clusters all reported similar feelings of institutional support.

Regarding career awareness, all three groups of students entered the learning community with very similar, moderate

levels. By the end of the year, all the clusters had substantial gains and reached high levels. As with the other domains, the ending levels of Cluster 1 and 3 were higher than those of Cluster 2, but unlike self-efficacy and sense of belonging, the career awareness of Cluster 2 finished higher than the starting point of Cluster 1.

Across the domains of research self-efficacy, sense of belonging, and career awareness, Clusters 2 and 3 both begin at very similar levels and experience statistically significant growth. However, Cluster 3 experienced much greater increases than Cluster 2 and reached parity with the higher scoring Cluster 1. Understanding the differential response of these two clusters may offer valuable insight into which students benefit the most from learning communities and why. The motivation and interest domain may be informative; unlike the other three domains, in this domain, Cluster 3 started at a high level, closer to Cluster 1, compared to the moderately scoring Cluster 2. Additionally, while Clusters 1 and 2 did not change, Cluster 3 increased and ended up as the highest scoring cluster.

Perhaps the higher level of motivation and Interest in Cluster 3 can explain its greater growth in the other domains than Cluster 2. Expectancy-value theory describes the nature of achievement motivation (Wigfield, 1994). This theory posits that “individuals’ expectancies for success and the value they have for succeeding are important determinants of their motivation to perform different achievement tasks” (Wigfield, 1994, p. 50). Expectancies were defined as individuals’ anticipations that their performance will either succeed or fail (Atkinson, 1957). Expectancies thus encompass individuals’ beliefs about how well they will do on a task (upcoming, currently, or in the future), as well as individuals’ beliefs about their current competence or ability (Eccles, 2005). Expectancies are associated with self-efficacy beliefs regarding individuals’ capabilities to accomplish a certain task and measures along with individuals’ confidence in how well they can perform the task (Schunk & Pajares, 2009). On the other hand, value can take any of the three forms: (a) attainment value refers to how an individual assigns importance to a particular task as that task is a representation of individuals’ identity (Cooper et al., 2017) and defines their competence. (b) Intrinsic value refers to the delight that an individual attains after accomplishing the task (Cooper et al., 2017). (c) The utility value or usefulness, on the other hand, refers to how a task fits into an individual’s plans (Wigfield & Cambria, 2010). Utility value has been related to motivation because by performing a useful activity, the activity itself becomes a means to an end, where such end can be the attainment of a certain occupation (Wigfield & Cambria, 2010).

Through the lens of Expectancy-value theory, perhaps the higher motivation and interest of Cluster 3 can be interpreted as greater utility value on the learning community. The similar initial sense of belonging scores argues against a difference in

attainment value between the two clusters. The public perception of statistics and data science as a popular career may direct students to perceive participation in this learning community in terms of future plans. The substantial growth in career awareness of all three clusters is indicative of the occupational focus of the learning community. Additionally, differential growth in the research self-efficacy subdomains of self-efficacy to pursue a research career, research communication, and research applications are potentially instructive. Cluster 3 had very large gains in these three categories and ended up surpassing the higher-achieving Cluster 1. These three subdomains are all applied career-focused skills, further reinforcing greater utility value for Cluster 3.

Other studies that have used Expectancy-value theory as a mechanism to study disciplinary communities have identified relevant results. For instance, in the context of introductory biology courses, achievement-related behavior has been reported as a joint function of individuals’ expectancy of success and the subjective value placed on such success (Sullins et al., 1995). Specifically, as might be expected, biology majors were found to place higher subjective value on success in the course than non-majors. Additionally, a subjective value significantly predicted students’ intent to enroll in future biology courses (Sullins et al., 1995). Similarly, a study of a virtual community also found that motivation significantly influences sustained participation intention (Sun et al., 2012).

Participation in linked classes as part of a STEM learning community has been reported to positively influence students’ attitudes, learning experiences, and motivation in STEM (Freeman et al., 2008). Indeed, the motivational benefits of a curricular focus on student interests are the reason *d'être* of discipline-based learning communities. Notably, the decline in *both* the motivation and interest subdomain scores for Cluster 2 stands out as the only subdomain scores that decline over the year. Meanwhile, in addition to a higher starting motivation and interest level, the students of Cluster 3 were unique in experiencing growth in this domain. Perhaps these trends point to feedback loops between motivation and interest, engagement in the discipline, and the other psychosocial domains. As the students of Cluster 2 learned more about the discipline of research in the context of statistics and data science, they may have grown less motivated and interested by the field, and therefore less engaged with the learning community resulting in more limited growth in the other psychosocial domains. In contrast, the students of Cluster 3 may have experienced a virtuous cycle wherein higher motivation and interest in the discipline lead to greater engagement and further growth in the other psychosocial domains. For example, the impressive growth by Cluster 3 in the aforementioned three applied research self-efficacy subdomains could reflect participation in extracurricular research or other activities by this cohort.

One major limitation of our study was the inability to match our survey results with student background data.

Could the more moderate prior research self-efficacy of Cluster 2 and 3 reflect differences in student preparedness entering the learning community? The fact that all three cohorts reported similar levels of career awareness argues against this possibility, but it remains an important avenue for further study. Similarly, data regarding demographic background or academic outcomes were not considered, thus reflecting our study's second limitation. It would have been very valuable to understand whether there were differences in academic preparation associated with the three student clusters. Additionally, it would have been useful to know the extent to which self-reported self-efficacy scores aligned with GPA, undergraduate research participation, and other independent metrics for student academic success.

### Implications for Learning

Our findings suggest that the alignment of the learning communities' curricular focus and the students' motivation and interest may be instrumental in resulting in positive psychosocial effects on students. Particularly for those who start with more moderate levels of self-efficacy and a sense of belonging. This study has implications for implementing learning communities that intentionally integrate the four components of cognitive apprenticeships. Cognitive Apprenticeship (Collins et al., 1991; Collins & Kapur, 2014) describes the components of learning environments that promote the development of cognitive and metacognitive skills by merging "the content being taught, the pedagogical methods employed, the sequencing of learning activities, and the sociology of learning" (Collins et al., 1989, p. 3). While learning environments often integrate (a) the content in the form of types of knowledge required for expertise, (b) the method regarding the learning strategies, pedagogical approach, or teaching methods used, and (c) the sequencing regarding the structure and the order of the tasks to optimize meaningful student engagement; less often a focus is placed on the sociology of learning. Sociology refers to the context within which learning experiences are situated via applying skills to realistic problems. Living learning communities place a stronger emphasis on the sociology of learning. The additional value of living together in the same residence hall allows for further synergies between teams and other faculty mentors, resulting in stronger self-reported learning outcomes (Inkelas et al., 2008) and moderate to strong psychosocial effects as shown in our study.

### Conclusion and Future Work

This study investigated the effects on interest and motivation, self-efficacy beliefs, sense of belongingness and levels of socialization, and career awareness of participating in a year-long residential research-oriented learning community. Our

analysis identified three groups before entering the learning community and their changes after the year-long experience. Similarities among the three groups at the end of the year-long experience were that all of the students ended with high levels of research self-efficacy beliefs, high levels of sense of belonging and socialization, and high levels of career awareness. Differences among the three groups at the end of the year-long experience related to their levels of motivation and interest, where one group reported high levels and stayed the same, a second group reported moderate levels and stayed the same, and the third group reported high levels and still experienced statistically significant increases in their levels of motivation and interest. These findings suggest that the intentional orchestration of content, method, sequencing, and sociology aspects of the learning community, irrespective of students levels of motivation and interest, had an impact on their research and data self-efficacy beliefs, sense of belongingness and socialization within the learning community, and career awareness in research-oriented fields.

Our study's limitation relates to the self-selection process for participation in the program and the moderate to high levels of self-efficacy beliefs and motivation and interest that students reported before starting the program. Consequently, a limitation of the study was the small sample size, which precluded the assessment of construct validity. However, we did validate our scale on the principles of content validity. The sample size limitation also resulted in two clusters (i.e., Cluster 2 and Cluster 3) with fewer students within each of them. Nevertheless, the clusters were indeed statistically significantly different from each other, as confirmed by a Welch's Test.

Similarly, since the sample was a mixture of students from five implementations of the learning community, each having the potential for changes, this might have reduced the reference value for the results. Despite the limitations of our study, and although studies reporting on institutional outcomes of learning communities in higher education are growing, studies focused on students' psychological factors are not common. Our study thus contributes to this literature by providing a specific example of a comprehensive learning community that integrated aspects of (a) shared coursework in the domain of statistics, probability, and data science, (b) a faculty-mentored, team-oriented data science research experience, (c) a living community in the same residence hall with social and academic activities, and (d) participation in a professional development seminar for career exploration in research-oriented fields; all providing a cognitive apprenticeship. Therefore, our future work will continue to investigate the characteristics of effective learning spaces that embed situated learning theory and communities of practice, along with their effectiveness on student learning, achievement, and attitudes.

**Data Availability Statement** Data available on request from the authors.

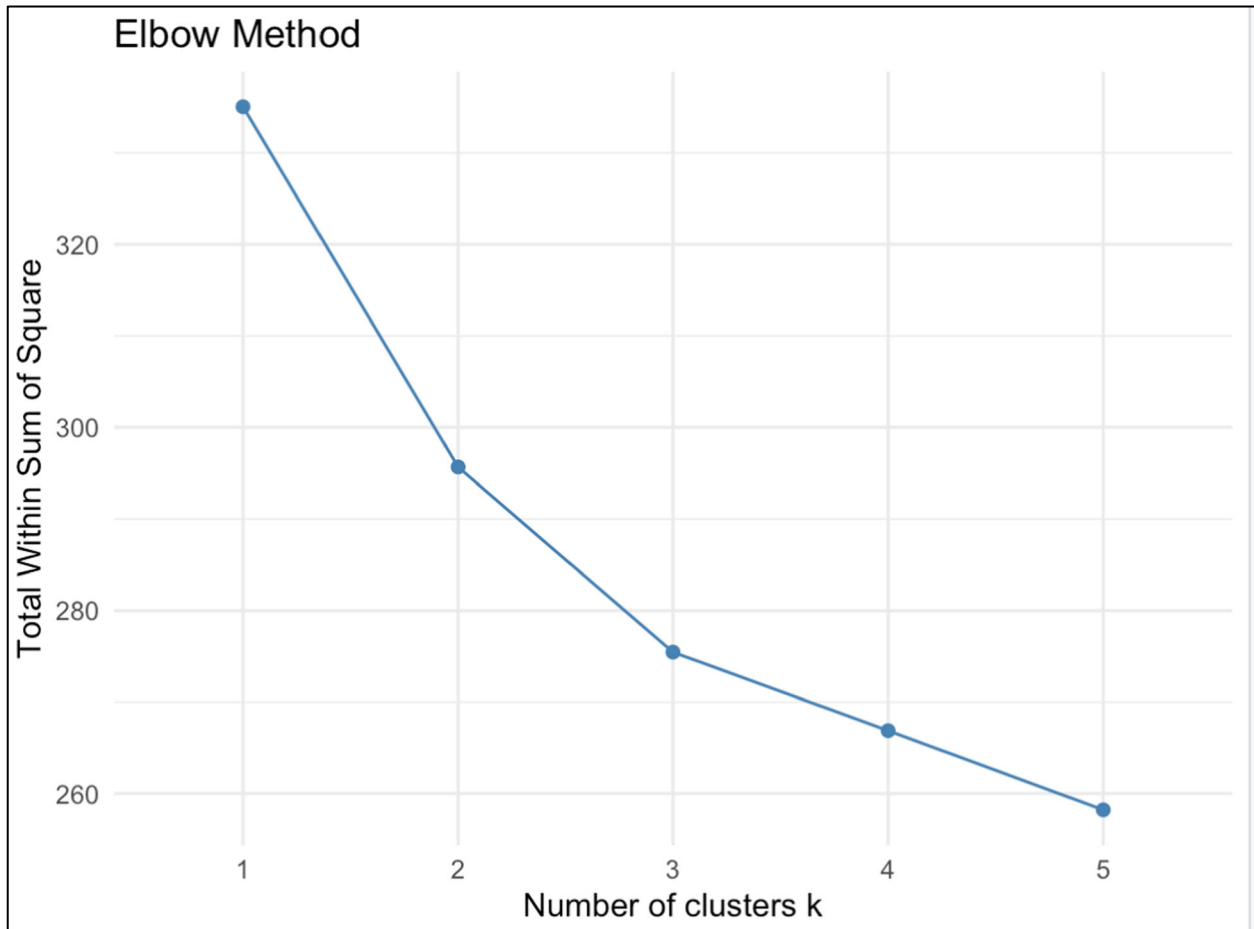
## Appendix

**Table 4** Statistics and Data Science Learning Community Student Survey (SDSLCSS)

Item	Question	Sub-Constructs	Domain/ Construct
Q1	I possess the motivation and persistence required for a career in a research-oriented field	Motivation	Motivation and Interest
Q2	My desire to become a researcher is strong enough to help me overcome most barriers I might encounter in pursuit of this career	Motivation	Motivation and Interest
Q3	I have a strong interest in pursuing a career as a researcher	Interest	Motivation and Interest
Q4	College faculty can encourage and promote my interest in pursuing a research-oriented field	Interest	Motivation and Interest
Q5	I have a desire to pursue graduate education	Interest	Motivation and Interest
Q6	I have an intention to choose a research-oriented career	Interest	Motivation and Interest
Q7	I have the abilities to become a successful researcher in future	Pursue a research career	Self-Efficacy
Q8	I have the ability to formulate a research question	Research Knowledge	Self-Efficacy
Q9	I have the ability to plan a research project	Research Knowledge	Self-Efficacy
Q10	I have the ability to conduct a research project	Research Knowledge	Self-Efficacy
Q11	I have the ability to understand the research process	Research Knowledge	Self-Efficacy
Q12	I have the ability to write a literature review	Research Knowledge	Self-Efficacy
Q13	I have the ability to understand a research paper or journal article	Research Knowledge	Self-Efficacy
Q14	I have the ability to document a research procedure	Research Knowledge	Self-Efficacy
Q15	I have the ability to write a research paper for publication	Research Knowledge	Self-Efficacy
Q16	I have the ability to prepare a research poster for presentation	Research Knowledge	Self-Efficacy
Q17	I have the ability to organize my research ideas in writing	Research Planning Skills	Self-Efficacy
Q18	I have the ability to work independently on a research project	Research Planning Skills	Self-Efficacy
Q19	I have the ability to conduct a search of literature for research purposes	Research Planning Skills	Self-Efficacy
Q20	I have the ability to observe and collect data	Data Analysis and Interpretation Skills	Self-Efficacy
Q21	I have the ability to organizing/entering data into a spread sheet	Data Analysis and Interpretation Skills	Self-Efficacy
Q22	I have the ability to statistically analyze data using computer software	Data Analysis and Interpretation Skills	Self-Efficacy
Q23	I have the ability to follow experimental or research procedures	Data Analysis and Interpretation Skills	Self-Efficacy
Q24	I have the ability to interpret data by relating results to the original hypothesis	Data Analysis and Interpretation Skills	Self-Efficacy
Q25	I have the ability to write the results of your experiment or research	Data Analysis and Interpretation Skills	Self-Efficacy
Q26	I have the ability to orally communicate the results of research projects	Data Communication skills	Self-Efficacy
Q27	I have an ability to understand the role of statistics in applied science	Data Communication skills	Self-Efficacy
Q28	I have the ability to apply what I learn in class to real world problems	Research Application skills	Self-Efficacy
Q29	I can work collaboratively with others	Groupwork skills	Self-Efficacy
Q30	I have opportunities to help out the local community	Groupwork skills	Self-Efficacy
Q31	I am comfortable interacting with people who are different from me	Groupwork skills	Self-Efficacy
Q32	I can work cooperatively and productively on a team	Groupwork skills	Self-Efficacy
Q33	I have the understanding of what graduate school may be like	NA	Career Awareness
Q34	I have the understanding of the various research-oriented career opportunities available to me	NA	Career Awareness
Q35	I have the understanding of research career options that I could specialize in	NA	Career Awareness
Q36	I feel a sense of belonging at [Blinded]	SOB: Institution	Sense of Belonging
Q37	I have opportunities to interact with [Blinded] faculty and staff	SOB: Institution	Sense of Belonging
Q38	I am comfortable talking to professors	SOB: Institution	Sense of Belonging
Q39	I have social support at [Blinded]	SOB: Institution Support	Sense of Belonging
Q40	It is easy for me to get to know other students with similar interests	SOB: with Others	Sense of Belonging
Q41	I participate in study groups	SOB: with Others	Sense of Belonging

\*SOB: Sense of Belonging

## Appendix 2 Output from Elbow Method for defining the number of clusters



**Table 5** Descriptive and inferential analysis of psychosocial effects for Cluster 1 over time

Psychosocial effect	T1		T2		Gain	Df	t	P-Val
	Mean	SD	Mean	SD				
Research self-efficacy beliefs	0.75	0.06	0.85	0.06	0.10	40	7.32	0.00
Pursue a research career	0.79	0.12	0.85	0.12	0.06	40	2.68	0.01
Research knowledge	0.67	0.10	0.80	0.08	0.13	40	6.27	0.00
Research planning skills	0.77	0.10	0.80	0.13	0.03	40	1.27	0.21
Data analysis/interpretation skills	0.76	0.10	0.90	0.07	0.15	40	6.91	0.00
Research communication skills	0.76	0.10	0.89	0.10	0.13	40	6.42	0.00
Research application skills	0.77	0.17	0.84	0.18	0.07	40	2.11	0.04
Group work skills	0.87	0.10	0.90	0.08	0.03	40	1.53	0.13
Motivation and interest	0.74	0.13	0.75	0.15	0.01	40	0.43	0.67
Motivation	0.75	0.12	0.75	0.16	-0.01	40	-0.23	0.82
Interest	0.74	0.15	0.75	0.16	0.02	40	0.72	0.48
Career awareness	0.56	0.19	0.87	0.12	0.31	40	10.16	0.00
Sense of belonging and socialization	0.83	0.11	0.89	0.10	0.05	40	3.69	0.00
SoB with institution	0.88	0.10	0.91	0.10	0.04	40	2.17	0.04
Institutional support	0.89	0.15	0.91	0.13	0.01	40	0.65	0.52
SoB with others	0.74	0.17	0.84	0.16	0.10	40	4.27	0.00

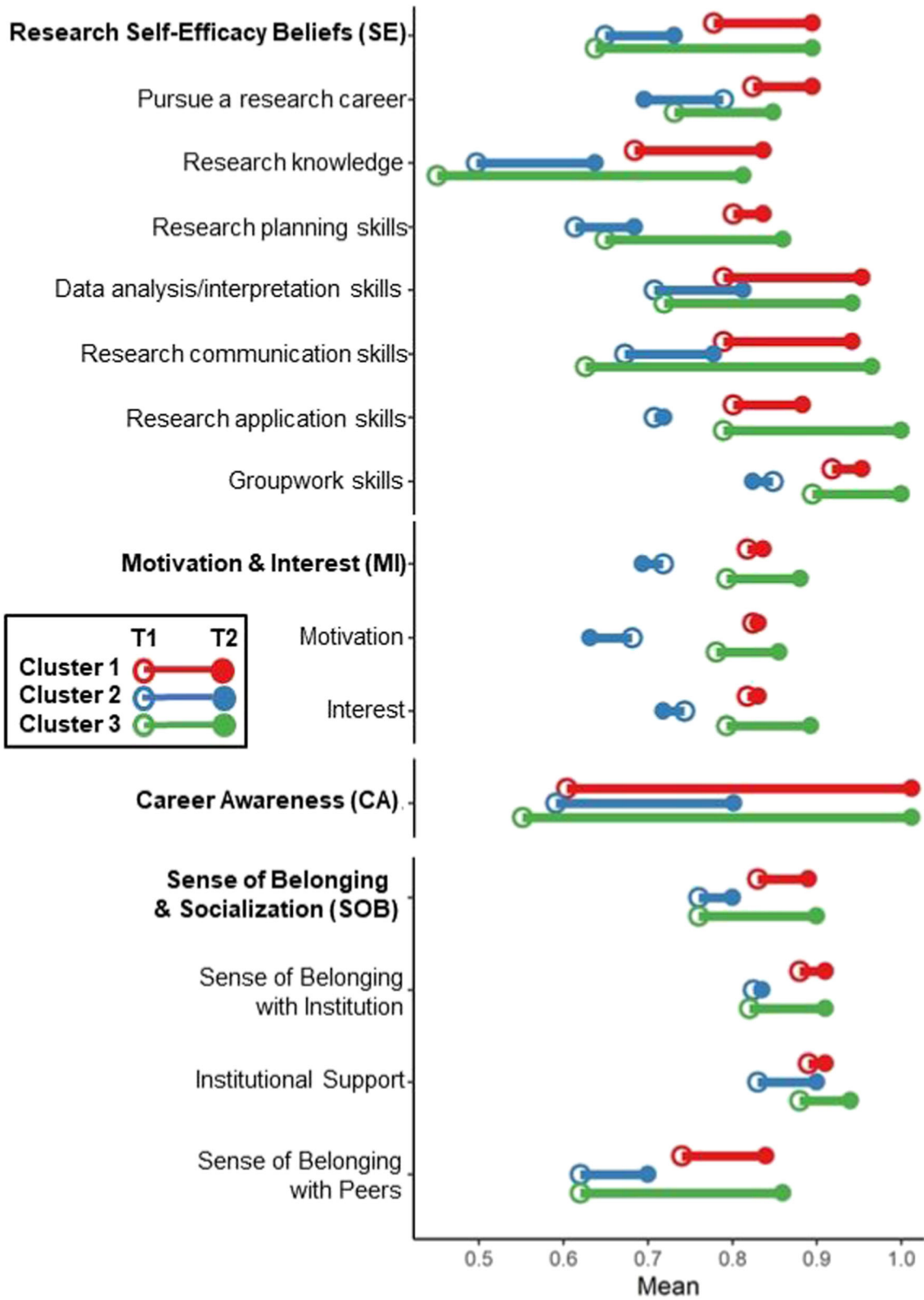
**Table 6** Descriptive and inferential analysis of psychosocial effects for Cluster 2 over time

Psychosocial effect	T1		T2		Gain	df	t	P-Val
	Mean	SD	Mean	SD				
Research self-efficacy beliefs	0.64	0.06	0.71	0.04	0.07	25	5.87	0.00
Pursue a research career	0.76	0.13	0.68	0.13	-0.08	25	-2.86	0.01
Research knowledge	0.51	0.07	0.63	0.05	0.12	25	9.61	0.00
Research planning skills	0.61	0.10	0.67	0.12	0.07	25	2.22	0.04
Data analy./interpretation skills	0.69	0.10	0.78	0.08	0.09	25	5.41	0.00
Research communication skills	0.66	0.13	0.75	0.12	0.09	25	3.19	0.00
Research application skills	0.69	0.20	0.70	0.15	0.01	25	0.21	0.83
Group work skills	0.81	0.07	0.79	0.11	-0.02	25	-1.01	0.32
Motivation and interest	0.66	0.13	0.64	0.16	-0.02	25	-1.01	0.32
Motivation	0.63	0.17	0.59	0.19	-0.04	25	-1.20	0.24
Interest	0.68	0.14	0.66	0.16	-0.02	25	-0.63	0.54
Career awareness	0.55	0.17	0.71	0.15	0.16	25	4.67	0.00
Sense of belonging and socialization	0.76	0.09	0.80	0.09	0.04	25	1.91	0.07
SoB with institution	0.83	0.10	0.83	0.13	0.00	25	0.00	1.00
Institutional support	0.83	0.16	0.90	0.12	0.07	25	2.21	0.04
SoB with others	0.62	0.12	0.70	0.11	0.08	25	3.03	0.01

**Table 7** Descriptive and inferential analysis of psychosocial effects for Cluster 3 over time

Psychosocial effect	T1		T2		Gain	df	t	P-Val
	Mean	SD	Mean	SD				
Research self-efficacy beliefs	0.63	0.04	0.85	0.06	0.22	16	11.42	0.00
Pursue a research career	0.71	0.13	0.81	0.14	0.10	16	2.38	0.03
Research knowledge	0.47	0.09	0.78	0.07	0.31	16	9.12	0.00
Research planning skills	0.64	0.12	0.82	0.08	0.18	16	5.60	0.00
Data analysis/interpretation skills	0.70	0.07	0.89	0.08	0.19	16	7.19	0.00
Research communication skills	0.62	0.10	0.91	0.09	0.28	16	8.44	0.00
Research application skills	0.76	0.13	0.94	0.09	0.18	16	5.22	0.00
Group work skills	0.85	0.09	0.94	0.08	0.09	16	4.51	0.00
Motivation and interest	0.72	0.15	0.79	0.18	0.07	16	2.60	0.02
Motivation	0.71	0.16	0.77	0.19	0.07	16	2.17	0.05
Interest	0.72	0.16	0.80	0.19	0.08	16	2.48	0.02
Career awareness	0.52	0.12	0.87	0.10	0.35	16	9.05	0.00
Sense of belonging and socialization	0.77	0.13	0.90	0.11	0.13	16	4.52	0.00
SoB with institution	0.82	0.15	0.91	0.10	0.09	16	3.73	0.00
Institutional support	0.88	0.14	0.94	0.14	0.06	16	1.43	0.17
SoB with others	0.62	0.19	0.86	0.15	0.24	16	4.44	0.00

**Appendix 6 Average scores in T1 and T2 regarding the four psychosocial effects.**



**Acknowledgements** The research reported in this paper was supported in part by the U.S. National Science Foundation under the awards DMS-1246818 and EEC-1449238 and by the Lilly Endowment Charting the Future Phase I Planning Grant, through the Purdue Office of the Provost. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. M.D. Ward's research is also supported by National Science Foundation (NSF) grants CCF-0939370, and OAC-2005632, by the Foundation for Food and Agriculture Research (FFAR) grant 534662, by the National Institute of Food and Agriculture (NIFA) grants 2019-67032-29077 and 2020-70003-32299, by the Society Of Actuaries grant 19111857, by Cummins Inc. grant 20067847, by Sandia National Laboratories grant 2207382, and by Gro Master.

## Declarations

**Conflict of Interest** The authors declare that there is no conflict of interests.

**Ethical Approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

## References

- Anthoine, E., Moret, L., Regnault, A., Sébille, V., & Hardouin, J.-B. (2014). Sample size used to validate a scale: A review of publications on newly-developed patient reported outcomes measures. *Health and Quality of Life Outcomes*, *12*(1), 1–10. <https://doi.org/10.1186/s12955-014-0176-2>.
- Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development*, *60*(3), 383–398. <https://doi.org/10.1007/s11423-012-9235-8>.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*(6, Pt.1), 359–372. <https://doi.org/10.1037/h0043445>.
- Baker, S., & Pomerantz, N. (2000). Impact of learning communities on retention at a Metropolitan University. *Journal of College Student Retention: Research, Theory & Practice*, *2*(2), 115–126. <https://doi.org/10.2190/62P5-CQ2U-NTUW-DM1C>.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>.
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachandran (Ed.), *encyclopedia of human behavior* (Vol. 4, pp. 71–81). New York: Academic press. (reprinted in H. Friedman [Ed.], *encyclopedia of mental health*. Academic Press, 1998).
- Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-Efficacy Beliefs of Adolescents*, *5*(1), 307–337.
- Barefoot, B. O. (2000). The first-year experience: Are we making it any better? *About Campus*, *4*(6), 12–18.
- Battaglia, O. R., Di Paola, B., & Fazio, C. (2015). Cluster analysis of educational data: An example of quantitative study on the answers to an open-ended questionnaire. *arXiv preprint arXiv:1512.08998*.
- Bielaczyc, K., & Collins, A. (1999). Learning communities in classrooms: Advancing knowledge for a lifetime. *NASSP Bulletin*, *83*(604), 4–10. <https://doi.org/10.1177/019263659908360402>.
- Blaney, J. M., & Stout, J. G. (2017). Examining the relationship between introductory computing course experiences, self-efficacy, and belonging among first-generation college women. *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, 69–74. <https://doi.org/10.1145/3017680.3017751>.
- Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 229–270). The MIT Press.
- Browne, M. N., & Minnick, K. J. (2005). The unnecessary tension between learning communities and intellectual growth. *College Student Journal*, *39*(4), 775–783.
- Carrino, S. S., & Gerace, W. J. (2016). Why STEM learning communities work: The development of psychosocial learning factors through social interaction. *Learning Communities: Research & Practice*, *4*(1), 3.
- Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator: The Professional Journal of the American Federation of Teachers*, *15*(3), 6–11 38–46.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Lawrence Erlbaum Associates, Inc.
- Collins, A., & Kapur, M. (2014). Cognitive apprenticeship. In R. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (Cambridge handbooks in psychology, pp. 109–127). Cambridge University Press. <https://doi.org/10.1017/CBO9781139519526.008>.
- Contreras, F. (2011). Strengthening the bridge to higher education for academically promising underrepresented students. *Journal of Advanced Academics*, *22*(3), 500–526. <https://doi.org/10.1177/1932202X1102200306>.
- Cooper, K. M., Ashley, M., & Brownell, S. E. (2017). Using expectancy value theory as a framework to reduce student resistance to active learning: A proof of concept. *Journal of Microbiology & Biology Education*, *18*(2), 1–8. <https://doi.org/10.1128/jmbe.v18i2.1289>.
- Crocker, L. (2001). Content validity. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 2702–2705). Elsevier.
- Dagley, M., Georgiopoulos, M., Reece, A., & Young, C. (2016). Increasing retention and graduation rates through a STEM learning community. *Journal of College Student Retention: Research, Theory & Practice*, *18*(2), 167–182. <https://doi.org/10.1177/1521025115584746>.
- Daly, C., Pachler, N., & Pickering, J. (2003). Teacher learning and computer-mediated communication. *International Journal of Learning*, *10*, 2897–2907.
- Dawson, S., Burnett, B., & O'Donohue, M. (2006). Learning communities: An untapped sustainable competitive advantage for higher education. *International Journal of Educational Management*, *20*(2), 127–139.
- DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, *39*(2), 155–164. <https://doi.org/10.1111/j.1547-5069.2007.00161.x>.
- Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics - Physics Education Research*, *5*(2), 020103. <https://doi.org/10.1103/PhysRevSTPER.5.020103>.
- Dodd, P. M. (2002). *Assessing the efficacy of learning communities at four North Texas community colleges*. PhD Dissertation, University of North Texas, Denton, TX, US.
- 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.



- Fink, J. E., & Inkelas, K. K. (2015). A history of learning communities within American higher education. *New Directions for Student Services*, 2015(149), 5–15. <https://doi.org/10.1002/ss.20113>.
- Franklin, K. (2000). Shared and connected learning in a freshman learning community. *Journal of The First-Year Experience & Students in Transition*, 12(2), 33–60.
- Freeman, K. E., Alston, S. T., & Winborne, D. G. (2008). Do learning communities enhance the quality of Students' learning and motivation in STEM? *The Journal of Negro Education*, 77(3), 227–240.
- Garcia, P. (1991). Summer bridge: Improving retention rates for under-prepared students. *Journal of The First-Year Experience & Students in Transition*, 3(2), 91–105.
- García -Valderrama, T., Mulero-Mendigorri, E. (2005). Content validation of a measure of R&D effectiveness. *R&D Management*, 35(3), 311–331. <https://doi.org/10.1111/j.1467-9310.2005.00392.x>
- Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools*, 30(1), 79–90.
- Holt, E. A., & Nielson, A. (2019). Learning communities and unlinked sections: A contrast of student backgrounds, student outcomes, and in-class experiences. *Research in Higher Education*, 60(5), 670–683.
- Inkelas, K. K., Soldner, M., Longbeam, S. D., & Leonard, J. B. (2008). Differences in student outcomes by types of living–learning programs: The development of an empirical typology. *Research in Higher Education*, 49(6), 495–512.
- Jessup-Anger, J. E. (2015). Theoretical foundations of learning communities. *New Directions for Student Services*, 2015(149), 17–27.
- Johnson, C. M. (2001). A survey of current research on online communities of practice. *The Internet and Higher Education*, 4(1), 45–60.
- Kogan, J. (2007). *Introduction to clustering large and high-dimensional data*. Cambridge University Press.
- Kuh, G. D. (2008). Excerpt from high-impact educational practices: What they are, who has access to them, and why they matter. *Association of American Colleges and Universities*, 14(3), 28–29.
- Lardner, E., & Malnarich, G. (2009). When faculty assess integrative learning. *Change: The Magazine of Higher Learning*, 41(5), 29–35.
- Lenning, O. T., and Ebbers, L. H. (1999). The powerful potential of learning communities: Improving education for the future. ASHE-ERIC higher education report, 26 (6), The George Washington University, Graduate School of Education and Human Development.
- MacGregor, J. (1991). What differences do learning communities make. *Washington center news*, 6(1), 4–9.
- MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, 35(2), 293–334.
- Medová, J., & Bakusová, J. (2019). Application of hierarchical cluster analysis in educational research: Distinguishing between transmissive and constructivist oriented mathematics teachers. *Statistika: Statistics and Economy Journal*, 2(2/2019), 142–150.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5–8.
- Parker, L. C., & Ward, M. D. (2019). Purdue University: Statistics living learning community. In T. Skipper (Ed.), *Aligning institutional support for student success: Case studies of sophomore-year initiatives*, National Resource Center for the first-year experience & students in transition. University of South Carolina.
- Peters-Burton, E. E., Merz, S. A., Ramirez, E. M., & Saroughi, M. (2015). The effect of cognitive apprenticeship-based professional development on teacher self-efficacy of science teaching, motivation, knowledge calibration, and perceptions of inquiry-based teaching. *Journal of Science Teacher Education*, 26(6), 525–548.
- Pike, G. R. (1999). The effects of residential learning communities and traditional residential living arrangements on educational gains during the first year of college. *Journal of College Student Development*, 40(3), 269–284.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. *Organizational Research Methods*, 19(2), 159–203.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497.
- Price, D. V. (2005, December). *Learning communities and student success in postsecondary education: A background paper*. MDRC. Retrieved from <https://files.eric.ed.gov/fulltext/ED489439.pdf>
- Schunk, D. H., Meece, J. R., & Pintrich, P. R. (2012). *Motivation in education: Theory, research, and applications*. Pearson Higher Ed.
- Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In K. R. Wenzel & A. Wigfield (Eds.), *Educational psychology handbook series. Handbook of motivation at school* (pp. 35–53). Routledge/Taylor & Francis Group.
- Solanki, S., McPartlan, P., Xu, D., & Sato, B. K. (2019). Success with EASE: Who benefits from a STEM learning community? *f*, 14(3), e0213827. <https://doi.org/10.1371/journal.pone.0213827>.
- Sullins, E. S., Hernandez, D., Fuller, C., & Tashiro, J. S. (1995). Predicting who will major in a science discipline: Expectancy-value theory as part of an ecological model for studying academic communities. *Journal of Research in Science Teaching*, 32(1), 99–119. <https://doi.org/10.1002/tea.3660320109>.
- Sun, Y., Fang, Y., & Lim, K. H. (2012). Understanding sustained participation in transactional virtual communities. *Decision Support Systems*, 53(1), 12–22. <https://doi.org/10.1016/j.dss.2011.10.006>.
- Taylor, K., Moore, W. S., MacGregor, J., & Lindblad, J. (2003). *Learning community research and assessment: What we know now (National Learning Communities Project Monograph Series)*. Washington Center, The Evergreen State College.
- Tinto, V., & Engstrom, C. (2008). Access without support is not opportunity. *Inside Higher Ed*, 9, 8–18.
- Tinto, V., & Goodsell, A. (1994). Freshman interest groups and the first-year experience: Constructing student communities in a large university. *Journal of The First-Year Experience & Students in Transition*, 6(1), 7–28.
- Tinto, V., & Russo, P. (1994). Coordinated studies programs: Their effect on student involvement at a community college. *Community College Review*, 22(2), 16–25.
- Tokuno, K., & Campbell, F. (1992). The freshman interest group program at the University of Washington: Effects on retention and scholarship. *Journal of The First-Year Experience & Students in Transition*, 4(1), 7–22.
- Ursachi, G., Horodnic, I. A., & Zait, A. (2015). How reliable are measurement scales? External factors with indirect influence on reliability estimators. *Procedia Economics and Finance*, 20(15), 679–686.
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82–96. <https://doi.org/10.1037/0022-3514.92.1.82>.
- Ward Jr., J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244.
- Wenger, E. (2011). Communities of practice: A brief introduction. Retrieved on February 19th, 2021, from <https://scholarsbank.uoregon.edu/xmlui/handle/1794/11736>
- Wenger, E., & Snyder, W. M. (2000). Communities of practice: The organizational frontier. *Harvard Business Review*, 78(1), 139–146.
- Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review*, 6(1), 49–78. <https://doi.org/10.1007/BF02209024>.

- Wigfield, A., & Cambria, J. (2010). Expectancy-value theory: Retrospective and prospective. *Advances in Motivation and Achievement, 16*(part a), 35-70.
- Williams, B., Onsmann, A., & Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*(3). <https://doi.org/10.33151/ajp.8.3.93>.
- Yuan, C., & Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. *J—Multidisciplinary Scientific Journal, 2*(2), 226–235. <https://doi.org/10.3390/j2020016>.
- Zeldin, A. L., Britner, S. L., & Pajares, F. (2008). A comparative study of the self-efficacy beliefs of successful men and women in mathematics, science, and technology careers. *Journal of Research in Science Teaching, 45*(9), 1036–1058. <https://doi.org/10.1002/tea.20195>.
- Zhao, C.-M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education, 45*(2), 115–138. <https://doi.org/10.1023/B:RIHE.0000015692.88534.de>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.