Three Decades as an NSF REU Site: Lessons and Recommendations

Authors:

Andrew L. McDevitt[1,2]

Manisha V. Patel[1]

Aaron M. Ellison[1]

Authors' addresses:

1. Harvard Forest, Harvard University
   324 North Main Street
   Petersham, MA 01366
2. Department of Integrative Biology, University of Colorado Denver
   Campus Box 171, P.O. Box 173364
   Denver, CO 80217-3364

Corresponding author: Aaron M. Ellison; aellison@fas.harvard.edu

Keywords: research experiences for undergraduates (REU), undergraduate research experience (URE), Harvard Forest, STEM, cultural-historical activity theory (CHAT),

Abstract

For the past 30 years, the NSF's Research Experiences for Undergraduates (REU) program has supported thousands of undergraduate researchers annually. Each REU site operates independently with regards to their research mission and structure, leading to a complex educational milieu distinct from traditional classrooms and labs. Overall, REU sites are perceived as highly formative experiences for developing researchers. However, even with improved assessment practices over the past decade, best practices for student learning and long-term impact are limited. To address this limitation, we recommend the use of cultural historical activity theory (CHAT) as a unifying framework to study these diverse programs. CHAT provides guidance for the collection of qualitative information which can help characterize REU programs in an educationally meaningful context. Adoption of CHAT by REU sites would improve dialogue among interdisciplinary programs. Such networks could further incentivize collaboration with discipline based education researchers and result in improved evidence-based practices.

Improving Evidence-based Practices for REU Programs

Since 1987, the US National Science Foundation (NSF) has supported undergraduate research experiences (UREs) for thousands of undergraduates each year through its Research Experiences for Undergraduates (REU) program (Box 1). At various times during these 30 years, NSF has developed assessment tools and examined assessment data across REU sites, but empirical evidence of impacts of these programs on further educational attainment or employment in science, technology, engineering, and mathematics (STEM) fields remains limited [1, 2]. Harvard University established its Summer Research Program in Ecology at the

Harvard Forest (HF-SRPE) in 1985; NSF has provided continuous core support for the HF-SRPE since 1989.

Through nearly three decades, HF-SRPE has participated in the development and implementation of NSF's vision for REU sites. Simultaneously, we have enhanced student experiences and improved short- and long-term effectiveness of our program by regularly measuring and reflecting on its success and failures, and integrating our assessment data with REU-wide evaluations. Based on our experiences, we here advocate for greater integration of educational theory, notably cultural-historical activity theory (CHAT), to increase the efficiency by which REU sites use assessment data to improve their own programs. Increased dialogue among programs and sharing of administrative procedures and data will strengthen educational research collaborations across various programs and disciplines.

History of REU programs

Undergraduate research experiences are widely accepted as a way to strengthen student preparation within scientific disciplines by providing authentic opportunities in research laboratories [3]. One of the first programs to support these experiences was the Undergraduate Research Participation Program (URP), which NSF began in 1958 and eliminated in 1982 because of cuts in its education budget [4]. In 1987, NSF initiated the Research Experiences for Undergraduates (REU) program, following the paid internship model of URP. Since then, NSF has been one of the largest supporters of undergraduate research programs, investing $1.12 billion in funding between 2002 and 2017 (Figure 1). In the early 2000s, the education research literature began to reveal the utility of undergraduate research programs for recruiting women [5, 6] and minority students into STEM fields [7-9]. In 2003, NSF aligned REU program goals with these new findings and outlined REU funding priorities in their yearly congressional budgets (Figure 1A). The implementation of the America COMPETES Act of 2010 (P.L. 111-478 §514) strengthened initiatives to reach diverse participants, especially from institutions where STEM research opportunities are limited. It also mandated the tracking of participants for STEM matriculation and employment for at least three years following graduation. Around the same time, the Biology REU Leadership Committee began using a common assessment tool, the Undergraduate Research Student Self-Assessment [10] (URSSA), to evaluate common goals and improve communication about BIO-REU programs [11].

REU Sites are Unique Educational Settings

Individual REU sites are defined by their intellectual themes and the community of researchers that they bring together. Although a primary goal of the REU program is to prepare undergraduate students for careers in STEM fields by provide scientific and engineering research opportunities, the design of these educational experiences depends on the goals and values of each site. Sites vary based on their personnel, infrastructure, their intellectual pursuits, and the student populations they serve.

For example, the Harvard Forest is a 2500-ha ecological field station located in rural Massachusetts. Our physical isolation poses many logistical challenges for students living and working together. Feedback from our participants over the past three decades has revealed the importance of physical infrastructure, organizational structure, and cultural norms on student perception of their educational experience in our program. Recognizing a heightened need for student support, we developed structured actives and resources for our students. Many of these are ancillary to what scientific researchers view as the core scientific mission of an REU program, but they are essential for developing the soft skills and personal connections important for future success in STEM fields [12, 13].

Adding to the variation among programs is the interdisciplinary nature of many of them. Individual scientific and engineering disciplines often have different educational priorities, knowledge requirements, skillsets, and practices. However, it cannot be assumed that the learning experiences of students within an interdisciplinary program can be evaluated using the same tools that are used in discipline-based classrooms and labs where students participate in similar activities.

Evaluating Program Goals

Evaluation of students and programs can be done quantitatively or qualitatively. At first glance, quantitative data seems to be simple to collect and analyze with common statistical techniques, and the results can be scaled up from small samples to large populations. However, the apparent simplicity of numerical results often masks many latent conceptual elements. Qualitative information is used to elicit meaning from the data and reveal structures underlying them.

Prior to the adoption of URSSA, each individual REU sites selected and managed its own assessment protocol. During this time, assessment consisted of single-site case studies [14], internally developed surveys [15], and participant surveys [16]. Qualitative data elicited insightful findings about student experiences, but were resource intensive to collect and did not represent all programs. Quantitative surveys created less of a burden on programs and were widely used [2], but the surveys often consisted of conceptually ambiguous responses and, because they were developed by individual programs for internal use, rarely validated, and were not comparable across programs [2, 15].

REU site program directors (PDs) routinely are scientists with strong backgrounds in basic scientific research but often lack experience in designing and implementing assessments of complex learning environments. Although intimately familiar with the challenges facing their programs, REU PDs not well-versed in discipline-based educational research run the risk of expending unnecessary effort trying to reinvent analytical wheels [2], incorrectly apply qualitative and quantitative techniques [17], or violate ethical guidelines of research involving human subjects. Additionally, the often ephemeral nature of REU site funding (many programs last < 5 yrs) combined with small sample sizes (averaging between 8–10 participants/yr) have made it difficult to establish broad claims about the effectiveness and impact of REU programs. Although we encourage increased collaboration and data sharing among REU sites, such collaborations should revolve around educationally meaningful characteristics such as program goals, mentorship models, professional development opportunities, and research cultures rather than research disciplines or types of institutions. The former also are likely to be more likely to engage successfully with discipline-based education researchers.

Trial and Error Approach to Science Education

In addition to the challenges of evaluating various aspects of REU programs, we have observed that many REU programs—and especially newly funded ones—go through many growing pains as new evaluative practices are implemented. Although reflection on these practices is important for each program, it leads to unnecessary and costly redundancy that could be avoided by the implementation of established best practices. A clearinghouse of such best practices is lacking, so we provide an example that we revisit throughout the paper.

At HF-SRPE, we rely on both formal (surveys and interview protocols) and informal (conversations with students) evaluations to assess and improve our program. These evaluations are used both to tweak parts of the program on the fly and to suggest future programmatic improvements. For example, conversations with students within the first three weeks of our 11-week program revealed that many the students lacked a clear understanding of why they were doing the “things their mentor told them to do,” despite perceived efforts by their mentors to communicate the “big picture” of research projects. This uncertainty created high levels of stress among the students, who only four weeks later had to write formal abstracts of their research projects and subsequently present them in a summer research symposium. In 2012, a few mentors, acting on their own, introduced their projects by asking their mentees to write short (1–2-page) research proposals outlining their summer research. These mentors used these proposals to evaluate the degree to which the students really understood the project, opened additional opportunities for discussions about the research between students and mentors, and resulted in substantial improvements in students’ performance while reducing their anxiety.

Given the success of this initial informal effort, all our students now write a two-page research proposal on their summer projects after two weeks of daily, intensive field, lab, and library work with their mentors. The proposal provides the students with a road map of their summer research, facilitates discussion between the mentors and students on the overall summer research, encourages them to read the scientific literature, and has them writing scientific prose from the first day of the program. The program coordinator and program director use the research proposal to identify and provide additional support to students that are struggling with their research.

After four years of using this method, we have obtained substantial feedback from students and mentors on its effectiveness. Although the research proposal itself provides a valuable guide for the summer research, once written it has been rarely re-visited by either students or mentors, even during preparation of the students' final research presentations. We now are considering asking students to revisit their research proposal and convert it into a longer written report that includes not only background and theory but also methods, results, and a brief discussion. Such a report not only will provide students with a tangible product of their summer experience and is a good exercise in scientific writing, but also represents an archivable, written record of their summer work that could be developed by the student and mentor into a more formal manuscript. Indeed, 15% of our students since 2001 have co-authored and published peer-reviewed papers with their mentors [14].

The use (or lack thereof) of statistics by the students in their final, oral presentations also revealed that their research proposal was not serving as a "living document" throughout the summer. We have always encouraged mentors and students to discuss, learn, and work with statistics at all stages of the summer research, from sampling or experimental design through data collection and analysis. However, we have seen that students rarely would broach statistical topics with their mentors. While students waited for mentors to initiate these discussion, mentors rarely mentioned statistics until data had been collected, organized, and were ready for analysis. Then, students would scramble anxiously in the last one or two weeks to learn and understand the statistical methods they needed to analyze their data. We used these observations and discussions with students to develop a multi-session workshop on the R statistical language [15].

Although we do not teach statistics per se, as different methods are needed for different projects, introducing R early in the program gets students thinking about data, gives them comfort with a software tool, and most importantly, empowers them to initiate discussions of statistical questions with their mentors. This workshop continually receives strongly positive reviews in the post-program survey.

Transferability of Best Practices

To apply the lessons from our previous example, it is important to identify the various stakeholders (student, mentors, program staff); contextualize the learning environments before, during, and after the implementation of new practices; and often change our own perspective about the meaning of “success”. In our example, we have viewed changes in practice through the eyes of the PD. Measures of success could differ if our focus was on research outcomes by mentoring faculty or on critically understanding student learning. Depending on the circumstances or goals of an REU site, it is likely to be necessary to draw upon a certain perspective of a practice to apply it to a new situation. To do so efficiently requires an agreed upon schema to document and share programmatic information.

The Need for a Common Theoretical Framework

The flexibility afforded to REU sites by NSF encourages innovative pedagogical approaches but also increases the heterogeneity across programs as compared to traditional classroom settings. Perhaps the largest methodological limitation for studying experiences and outcomes of undergraduate researchers is that our assessment tools were developed without any overarching theoretical framework: an explicit statement of our theoretical assumptions with respect to learning. Rather, our surveys were developed with specific programmatic goals in mind. Although the surveys served their intended purpose, the lack of theory and abstraction limits our ability to relate our findings to the broader literature on education and outcomes or to compare them meaningfully to data from other REU sites. We experienced this same limitation when examining 10 years of HF-SRPE pre-post surveys [14]. Although we could predict difference in various learning gains based upon prior experiences, we had limited ability to explain the phenomena we observed. The design of our short self-reporting survey was an intentional compromise between sample size and survey depth, and such concessions contribute to a lack a general mechanistic understanding of outcomes of UREs [2].

Cultural Historical Activity Theory (CHAT) as a Broad Theoretical Framework

To unify the study of UREs, we propose using the meta-theoretical framework of Cultural Historical Activity Theory (CHAT). CHAT allows researchers to study the social construction of knowledge by examining both individual and collaborative activities [16]. CHAT uses a systems perspective to provide a blueprint describing the components that influence the social construction of knowledge [17]. This systems approach to research has proven useful for studying complex and heterogeneous educational phenomena and helps to derive meaning from seemingly contradictory information [18].

Originated by Vygotsky in the 1930's, the current "third generation" of CHAT activity systems are best visualized through what are known as activity triangles [16]. CHAT requires the acknowledgement of seven distinct elements ("nodes") that take part in an activity within a system of interest and the examination of connections ("edges") between them [16] (Figure 2):

1. *Subject* – The individual or group of focus during the specified activity;

2. *Object* – The goal or motive behind the specified activity;
3. *Rules* – The stated or unstated rules that govern how individuals act within the context of the specified activity;
4. *Community* – The social context in which the specified activity is conducted;
5. *Division of labor* – How tasks are shared among the community to accomplish the specified activity;
6. *Mediating artifacts* – The tools used in creating the *Object*;
7. *Outcome* – The effect generated by subject working in concordance with other components of the activity system to accomplish the *Object*.

The nodes of the triangles represent these elements and the edges represent interactions between them [17, 19].

REU programs are complex social learning environments and CHAT also provides the ability to make sense of contradictory information that arises within the system and through time [17]. These contradictions, which often are difficult to rationalize using other frameworks, are classified into four types: *Primary contradictions* (1°) exist within an element *(e.g.*, contradictory *Rules*); *Secondary contradictions* (2°) exist within interactions between two elements (*e.g.*, *Division of labor* is not aligned with *Mediating artifacts*); *Tertiary contradictions* (3°) are contradictions manifested during temporal transitions of an activity system (*e.g.*, mentors refining their approach during the program); *Quaternary contradictions* (4°) exist between similar activity systems of which the subject is a member (*e.g.*, REU experience compared to scientific coursework) [20].

CHAT can help make sense of the complex social learning environment across the many levels of organization characteristic of REU programs. Part of the reason evidence supporting the benefit of undergraduate research experiences is limited results from the difficulty in defining similarities and differences among different programs [2]. Small sample sizes within programs (averaging 8-10 participants) and ephemeral funding also make it difficult to establish lasting partnerships with education researchers whose experience in survey instrument selection and evaluative techniques that would benefit classically trained scientists. By characterizing programs using CHAT, we can envision examining system components of UREs and provide a platform for education researchers to connect with multiple program directors whose programs share common characteristics.

Values of the program experience

Primary contradictions within the CHAT framework are often a result of differing value judgements that underlie the system [20]. These contradictions are fundamental to the system and form the foundation of higher orders of contradictions [20, 21]. One primary contradiction that we see in REU programs is the balance between the participant's role as a learner of the nature of scientific practice and her role as an independent researcher. Vygotsky's "zone of proximal development" (ZPD) highlights this contradiction: it is the zone between an individual's ability to complete a task unaided and an individual's inability to complete a task even with assistance [22]. The ZDP is the middle ground in which tasks can be completed through social guidance and scaffolding [22]. To conduct research successfully, an independent researcher must master specific sets of skills, including hypothesis generation, problem solving, analytical techniques, and scientific writing and communication. Participating in mentored research provides students with the opportunity to learn from independent researchers while working towards a common research goal. Assigning tasks too far beyond the capacity of the participant inhibits learning from occurring and prevents research from progressing. Assigning tasks that a participant can accomplish unaided promotes research productivity, but assigning too many such tasks limits the student's potential to learn. Designing a program that facilitates the transition from novice to independent researcher while also generating research products requires a careful balance between tasks within the participant's current capacity and the ZPD.

Alignment of system components

After program values are established, components within an activity should complement one another to promote the *Object* and *Outcomes*. As is often the case, the structure of components may not be aligned. Secondary contradictions help to illuminate this misalignment and may lead to subsequent changes within the activity system [20]. For example, subjects may not have sufficient skillsets to interact with *Mediating artifacts*, the *Division of labor* may not be aligned with the present *Community*, or program *Rules* and expectations may conflict with the personal values of the *Subject*. Conflicts between system components result in specific obstacles that are manifestations of fundamental tensions (primary contradictions) within the activity system [21]. Because of this origin, it is often best to address the source(s) of the conflict rather than simply the symptoms.

Introduction of a new practice

To help resolve secondary contradictions and address underlying primary contradictions, change must occur in the activity system. Tertiary contradictions are differences in the system that occur at temporal transitions [20], and program directors may be interested in examining them as they change various instructional activities or procedures. As new procedures are implemented, a transition to the more "advanced practice" may not be immediate [20, 21].

Understanding resistance to change may reveal additional information about primary contradictions and potentially lead to smoother tertiary transitions. Revisiting our previous example of the research proposal provides insight into how examining program change through the lens of CHAT might improve the adoption of intervention techniques. Although the research proposal was a *Mediating artifact* that was meant to increase communication between students and members of their research team, it was not acting as a living document that would be revisited throughout the summer. Adoption seemed to be limited by the *Rules*, scaffolding, and additional *Mediating artifacts*. For example, during the first few iterations, not all students experienced the same benefits from their research proposals. Students with more exploratory projects found themselves reworking their methods throughout the summer; their final product did not reflect their proposal. For these individuals, the *Rules* of their project and the creating of the proposal were misaligned. Additionally, after the first two weeks, there were no additional programmatic *Mediating artifacts* that connected research projects back to the proposal.

Had we used CHAT, we might have been able to identify sooner how the design of the proposals were misaligned with the foci of some research projects. However, lessons from the proposal were applied to the creation of our R workshop. We recognized the need to design a *Mediating artifact* that was commensurate with the skill levels of our students and aligned with the needs of their projects. Because the workshop was spread out over four sessions, students could continue to revisit the concepts at different stages of project development.

Levels of examination within an REU program

Our examples have focused primarily on student-level activity systems, but the same framework also can be applied to examine mentor and program levels (Figure 1). It can act as the focus of primary through tertiary contradictions, and can illuminate neighboring activity systems at the student level. In CHAT, the latter, quaternary contradictions arise between adjacent activity systems, often triggered by tertiary contradictions [20, 21]. Transitions resulting in tertiary contradictions, such as we saw in the research proposal example, also may create disturbances between activity systems and some or all its neighbors. In the case of REUs, mentor- and program-level systems share the same *Object* but may differ in the structure of other components (Table I). Not only could the interactions between activity systems change, so too could the adjacent systems if there is enough overlap in system components.

Applying the CHAT framework to REU research

Recruitment and Hiring Practices

At one of the earliest stages of the program, a *primary contradiction* exists when we are selecting students for the program. We try to strike a balance between selecting students who appear to be best qualified (*i.e.*, most experienced) to conduct research and students who have the most to gain out of the experience. This conflict arises in part due to cultural biases of academic research in which success is measured through productivity. However, as mentors and educators, we also want to work with students who are willing to push beyond their comfort zone and maximize the impact of a research experience. At HF-SRPE, this *primary contradiction* is further complicated by the different stakeholders involved in the hiring process. Mentors advocate for their projects, funders push for students from certain institutions, demographics, academic majors, or skillsets, and program directors seek a lasting and cohesive identity for the program. Characterizing these various components and assessing if recruitment and hiring goals are being met is difficult, especially when hundreds or even thousands of applications are reviewed in a scant few weeks.

Demographic data and quantifiable metrics such as GPA may be the easiest variables to gather and compare over time or across programs. However, for REU programs to evaluate the effectiveness of practices such as recruitment, application requirements, and selection criteria, the data collected need to be aligned with the context. This is where a meta-theoretical framework like CHAT can be used to identify and prioritize useful data to be collected (Figure 3). Although it would be best to collect lines of evidence supporting each CHAT component, we would minimally expect to collect more informative characteristics about the individual (*Subject*), the priorities of the position (*Object*), the expectations of the hiring process (*Rules*), practices implemented in recruiting or selection (*Mediating artifacts*), and the hiring decision (*Outcome*). Rather than serving as a predictive theory about the hiring process (which could also be integrated into a research program), CHAT provides focus and clarity for understanding this complex system.

Understanding variation in learning gains

To make sense of assessments of learning gains, it is important to understand the educational contexts in which those gains were observed. Most concept inventories lack these contextual questions and require the education researcher to integrate this into the research design. Although concept inventories often are praised for their extensive reliability and validity testing, they usually are designed to assess traditional coursework and are rarely aligned with learning objectives of REU programs. For example, URSSA was developed to better capture the learning gains associated with undergraduate research experiences such as the REU program

[10]. However, criticisms of URSSA are that information is self-reported and measure mainly affective domains, (attitudes, feelings, motivation, *etc*.) rather than a direct assessment of knowledge.

If components of URSSA do align with specific research questions, CHAT can be used to inform data collection. One advantage of URSSA for the REU community is that supplemental questions are built into the survey platform in addition to the core questions related to learning gains. However, these questions, such as the demographics seen in Figure 4, rarely explain much of the variation in these gains nor are they consistently administered across programs. Although it would be ideal to align and characterize all seven components of the activity system with respect to learning gains (Figure 4), we hypothesize, at least for participants in HF-SRPE, that characterizing the skills and knowledge a student brings with them to the research experience (*Subject*), the resources available to the student during their research experience (*Mediating artifacts*), the level of support they received (*Division of labor*), and what success means given a student's research experience (*Object*) are likely going to be the most factors in explaining the variation in student learning gains.

Assessing the Impact of REU Programs

Feedback from previous REU participants suggest that mentored independent research is a formative experience for career development. Systematic, post-program tracking of REU participants is uncommon despite it being a legal requirement since 2010 (P.L. 111-478 §514). We annually survey past participants of HF-SRPE and data provide some support for long-term persistence and high rates of employment by HF-SRPE alumnae/i in STEM fields (Figure 5). However, we cannot account accurately for the distribution of non-responses, nor are we able to determine specific effects of HF-SRPE on individual decisions to pursue STEM careers. Like many of our colleagues who work with REU students, we believe that mentored research experiences launch them into STEM careers, but we cannot predict where they would be without this experience. Logistical and ethical constraints prevent researchers from forming true control groups and require the use of quasi-experimental designs. Expansions of data collection efforts before and during REU programs would help to characterize students and their experiences.

Again, the systems approach of CHAT can be used to identify important aspects of the complex learning environment and overcome the limitations of other study designs. For example, we would predict that factors such as the skills and knowledge a student brings with them to the research program (*Subject*), the professional development opportunities available to them during their research experience (*Mediating artifacts*), the interactions students have with other members of the research community (*Community*), and the goal of the research experience (*Object*) could be very influential on the long-term persistence in STEM disciplines and careers (*Outcome*).

Future Directions

As evaluative research continues to develop within the REU community, we see CHAT as a guideline for programs to follow. It can help to identify and coordinate the type of information that should be collected so that knowledge can be transferred more easily between programs. Without CHAT, or a similar theoretical framework, we will continue to be limited in the extent to which we can transfer meaningful experiences and best practices to our colleagues in the larger REU community. Although there is certainly value to developing and implementing new interventions within individual programs, the often short-term nature of REU sites highlights the need to evaluate what works and what doesn’t, and disseminate this information to the broader scientific community lest we keep reinventing pedagogical wheels. Admittedly, not everything can be or should be quantified. The heterogeneity between REU programs provides many challenges for science-education researchers to study. CHAT should encourage more researchers to collaborate with other programs.

Glossary

Affective domain: One of the three Bloom's Taxonomy Learning Domains. Unlike the cognitive (knowledge-based) and psychomotor (action-based) domains, the affective domain examines the development of emotional characteristics such as attitudes, feelings, motivation, and values.

CHAT: Cultural-historical activity theory is a systems approach to understanding complex human phenomina by understanding conflicts between individual and cultural values. In an educational setting, this can help to frame the social construction of knowledge.

Construct: An abstraction that helps to conceptualize phenomena that cannot be directly observed.

Concept Inventory: A rigorously designed test (often multiple chioce) that evaluates an individual's mastery of a particular topic. Incorrect responses often reflect common misconceptions and can be used to further distinguish the level of mastery of a topic.

Evidence-based Practice: Use of the current best evidence to imform pedigogical decisions.

Instructional intervention: An intentional program aimed at promoting a specific congnitive or affective goal.

Instrument: A scientifically validated survey or protocol for obtaing information from participants.

HF-SRPE: The Harvard Forest Summer Program in Ecology has received continued NSF REU funding since 1989. All authors have experience running this program and are providing insights and reccomendations based on this experience: Ellison has been the Program Director since

2004, Patel has been the Program Coordinator since 2012, and McDevitt was an on-site proctor and evaluation-researcher in 2015 and 2016.

REU: NSF's Research Experiences for Undergraduates program funds sites (such as HF-SRPE) and individual students to participate in STEM research oppertunities.

Scaffolding: Structured education activites that allow student to progress towards deeper understanding and independence in the learning process.

URE: A generic acronym for undergraduate research experiences whereas REUs are NSF's branded programs.

URSSA: The Undergraduate Research Student Self-Assessment is an evaluation tool for REU programs whose use is required by NSF.

ZDP: The zone of proximate development is the area between a student can learn with help and without help.

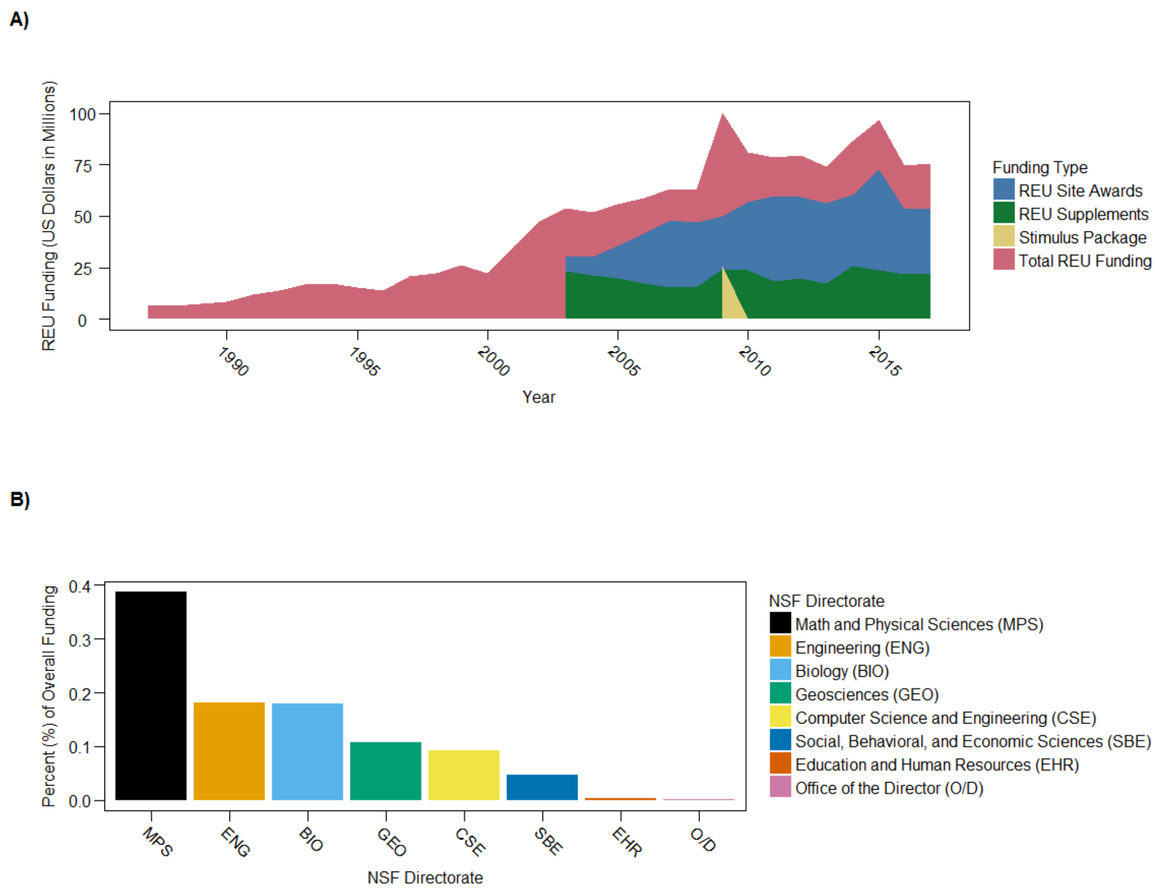

Figure 1. Funding for Research Experiences for Undergraduates (REU) programs. Support for REU programs based on A) yearly congressional allocations and B) NSF directorate support. Funding data (2002-2017) was compiled based on yearly NSF congressional budget requests. Archives of REU awards (nsf.gov/awardsearch/) provided estimates for remaining years and directorate contributions.

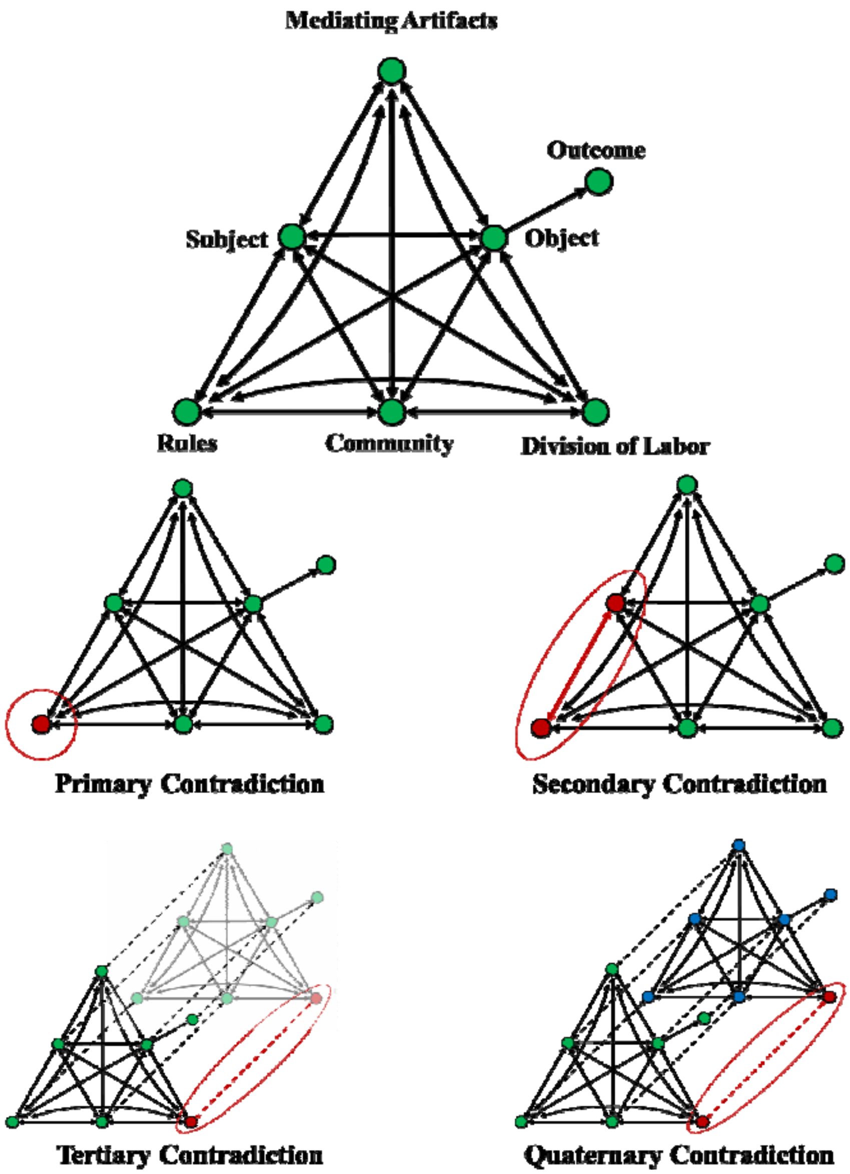

Figure 2. CHAT's activity system components. The activity triangle highlights A) how components interact with others within the system, and B) the contradictions that can be examined through CHAT.

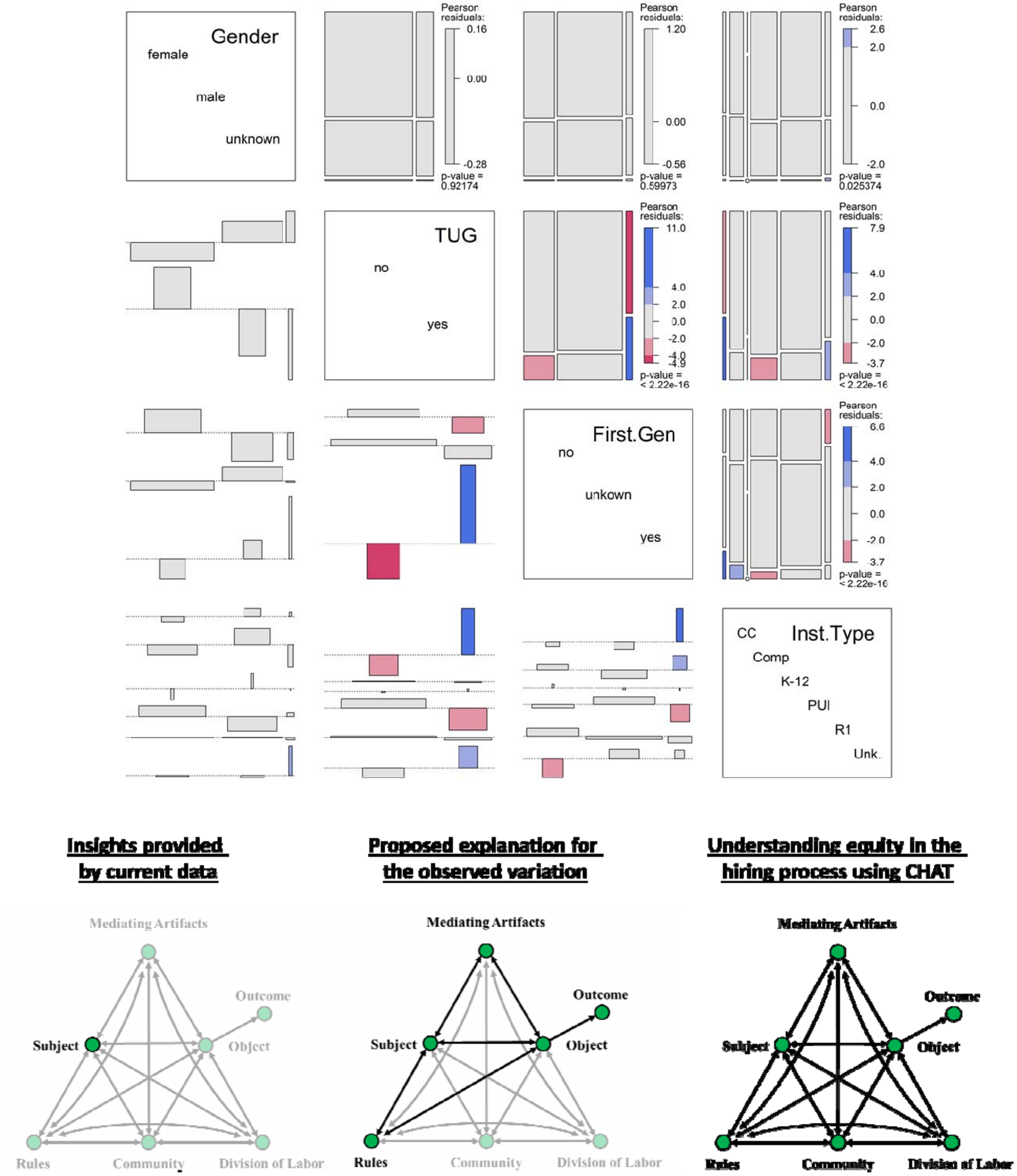

Figure 3. Commonly collected data during the recruitment and hiring process. The top panels show mosaic plots illustrating relative proportions of students in different groups of interest to students, program directors, programs, and funders. Significant positive (blue) and negative (red)

pairwise correlations are indicated. TUG: Student from a group traditionally under-represented in science; First.Gen: Student who is the first in their family to attend college or university; Inst.Type: type of institution, including community college (CC), Comprehensive university (Comp), K-12 (kindergarten through high school), PUI (primarily undergraduate institution), R1 (research-1 university), and Unk (unknown or not applicable). The CHAT activity triangles illustrate how components are assessed with current frameworks (left) or could be assessed within a full CHAT framework.

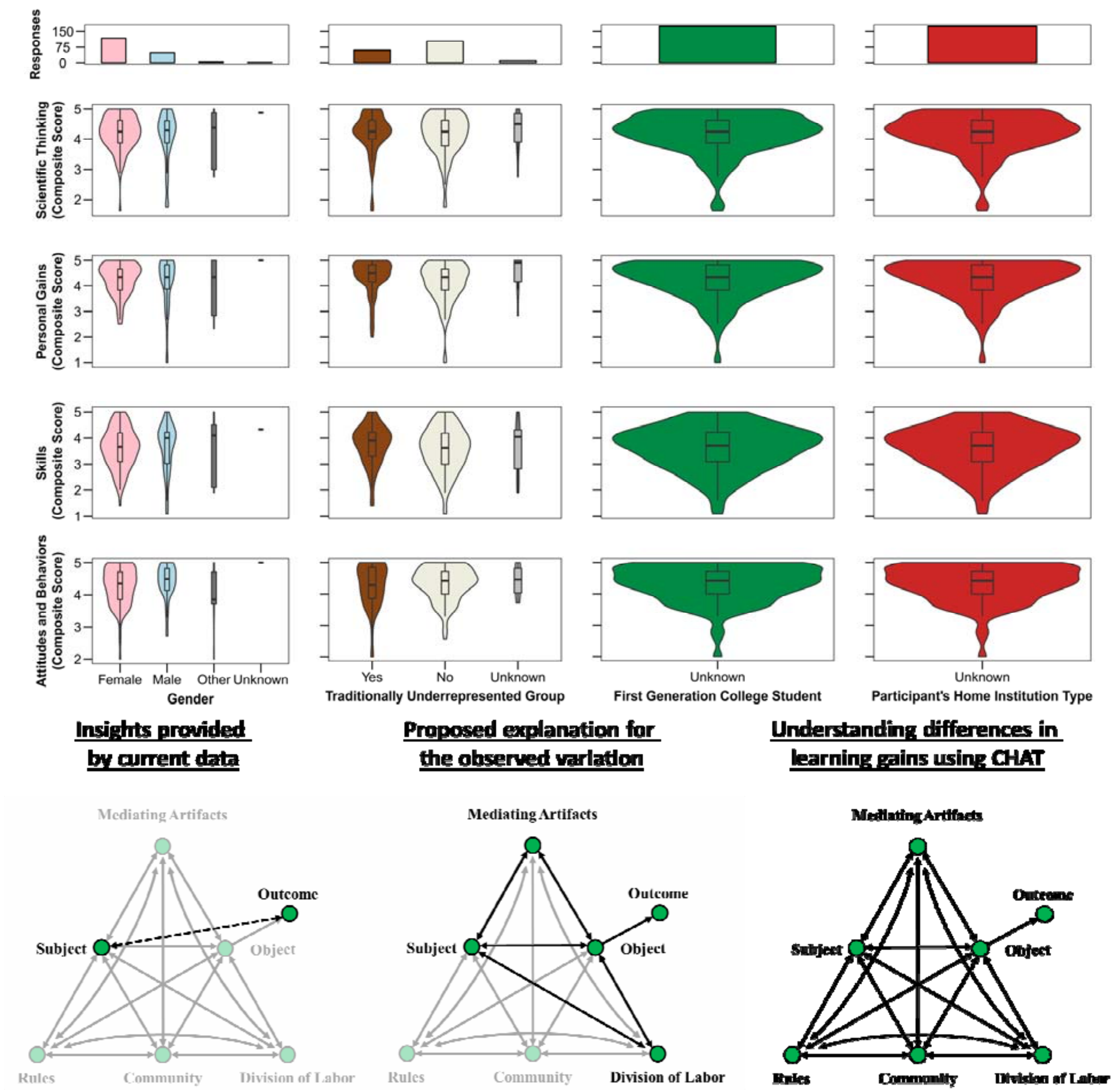

Figure 4. Commonly collected data when assessing learning gains. The top panels show changes in scientific thinking, personal gains in overall confidence in doing research, research skills, and attitudes and behaviors about doing research. Values range from 1 (low) to 5 (high) for all variables. The total number of participants in the different groups are shown in the top row; in the other panels, violin plots show the distribution of the data with inset box plots illustrating median, quartile, and upper and lower deciles of the data. More detailed analysis of these data,

collected from pre/post surveys given annually to HF-SRPE students are presented in [13]. The CHAT activity triangles illustrate how components are assessed with current frameworks (left) or could be assessed within a full CHAT framework.

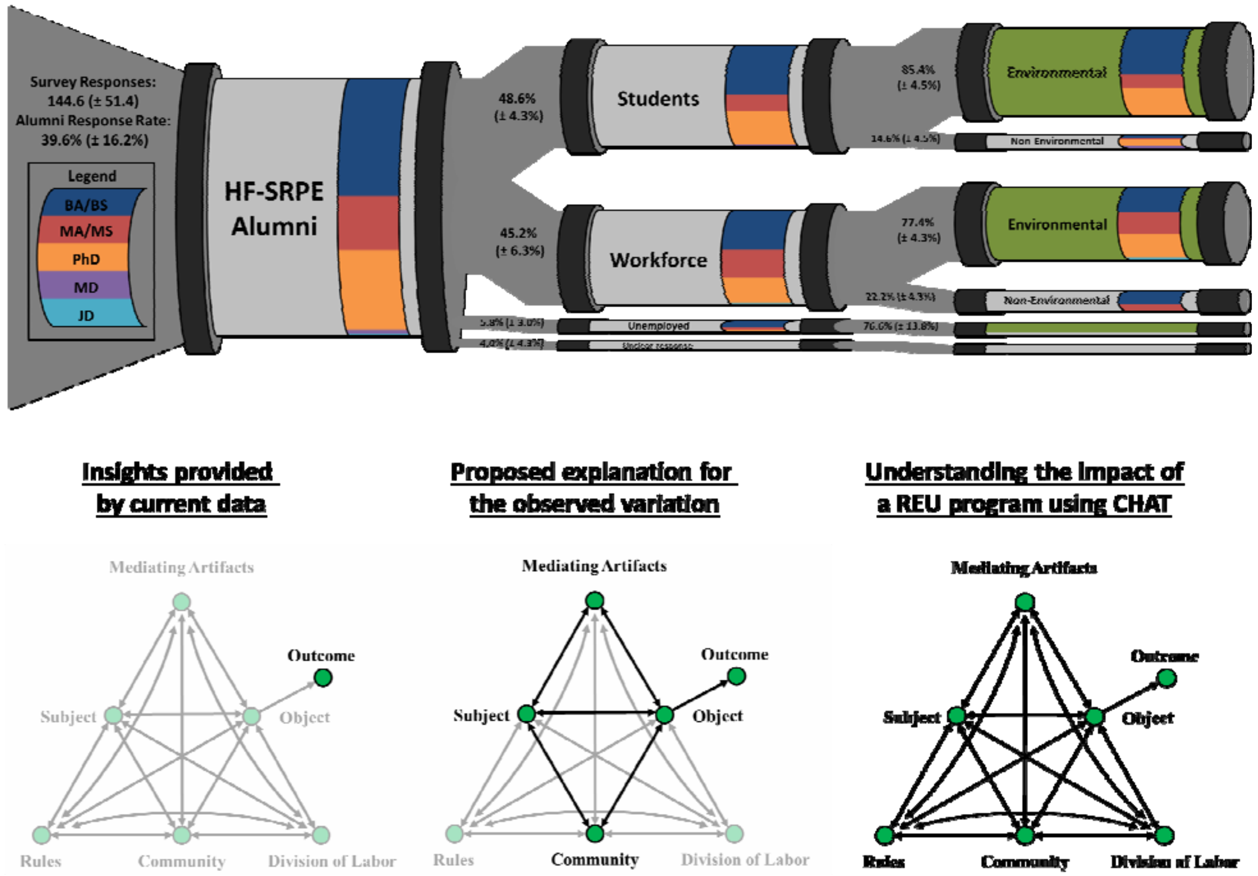

Figure 5. HF-SRPE career outcomes. Annual alumni surveys were sent to alumni (cohorts from 2001 onward) between 2012-2016. Averages of yearly snapshots reveal that alumni peruse/receive environmental/ecological related graduate degrees and continue to use these disciplines during their careers. Further information is required to determine the level of impact HF-SRPE had on these outcomes. The CHAT activity triangles illustrate how components are assessed with current frameworks (left) or could be assessed within a full CHAT framework.

References

1. Beninson, L.A. et al. (2011) Evaluation of the Research Experiences for Undergraduates (REU) Sites Program. CUR Quarterly 32 (1), 43-48.
2. Linn, M.C. et al. (2015) Undergraduate research experiences: Impacts and opportunities. Science 347 (6222), 1261757.
3. Kuh, G.D., High-impact educational practices: What they are, who has access to them, and why they matter, Association of American Colleges and Universities, Washington, DC, 2008.
4. Neckers, D. (1982) The threat to undergraduate research. J. Chem. Educ 59 (4), 329.
5. Kardash, C.M. (2000) Evaluation of undergraduate research experience: Perceptions of undergraduate interns and their faculty mentors. Journal of educational psychology 92 (1), 191.
6. Liang, B. et al. (2002) Mentoring College-Age Women: A Relational Approach. American Journal of Community Psychology 30 (2), 271-288.
7. Gregerman, S.R. et al. (1998) Undergraduate student-faculty research partnerships affect student retention. The Review of Higher Education 22 (1), 55-72.
8. Foertsch, J. et al. (2000) Summer research opportunity programs (SROPs) for minority undergraduates: A longitudinal study of program outcomes, 1986-1996. Council of Undergraduate Research Quarterly 20 (3), 114-119.
9. Nnadozie, E. et al. (2000) Undergraduate Research Internships and Graduate School Success.
10. Hunter, A.B. et al. (2009) URSSA: Evaluating student gains from undergraduate research in the sciences. CUR Quarterly 29 (3), 15-19.

11. BIO REU PI Workshop 2010, in: Christian, A. and Hannigan, R. (Eds.) BIO REU Leadership Committee, Arlington, VA, 2010.

12. National Research Council. (2013) Education for life and work: Developing transferable knowledge and skills in the 21st century, National Academies Press.

13. Kumar, S. and Hsiao, J.K. (2007) Engineers learn "soft skills the hard way": Planting a seed of leadership in engineering classes. Leadership and Management in Engineering 7 (1), 18-23.

14. McDevitt, A.M. et al. (2016) Insights into student gains from undergraduate research using pre-and post-assessments. BioScience 66 (12), 1070-1078.

15. R Core Team. (2016) R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing.

16. Roth, W.M. and Lee, Y.J. (2007) "Vygotsky's neglected legacy": Cultural-historical activity theory. Review of educational research 77 (2), 186-232.

17. Cole, M. and Engeström, Y. (1993) A cultural-historical approach to distributed cognition. Distributed cognitions: Psychological and educational considerations, 1-46.

18. Talbot, R. et al., Theoretically Framing a Complex Phenomenon: Student Success in Large Enrollment Active Learning Courses, Physics Education Research Conference Proceedings, 2016, pp. 344-347.

19. Yamagata-Lynch, L.C. (2010) Understanding cultural historical activity theory, Springer.

20. Engestrom, Y. (1987) Learning by expanding: An activity-theoretical approach to developmental research, Orienta-Konsultit.

21. Foot, K. and Groleau, C. (2011) Contradictions, transitions, and materiality in organizing processes: An activity theory perspective. First Monday 16 (6).

22. Vygotsky, L. (1987) Zone of proximal development. Mind in society: The development of higher psychological processes 5291, 157.