TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 00 (2021) 1–23

ISSN: 1756-8765 online
DOI: 10.1111/tops.12539

This article is part of the topic "2019 Rumelhart Prize Honoring Michelene T. H. Chi," Wayne D. Gray (Topic Editor). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview.

Translating a Theory of Active Learning: An Attempt to Close the Research-Practice Gap in Education

Michelene T. H. Chi

Mary Lou Fulton Teachers' College, Arizona State University

Received 6 October 2020; received in revised form 28 April 2021; accepted 29 April 2021

Abstract

Despite decades of research related to teaching and learning, the findings have made little impact on classroom teaching and learning. This paper briefly describes the four existing methods to close this gap, with more extensive analyses of the limitations of one of the four methods, which is to consolidate and distill robust laboratory findings reported over the past decades and attempt to translate them for classroom practice. An alternative method is proposed, which is to translate a theory of how students learn, called *Interactive, Constructive, Active, Passive* (ICAP), so that teachers and practitioners can translate their understanding of such a theory into practice themselves, thereby giving teachers autonomy, flexibility, generalizability, and ownership of their own designed interventions based on ICAP. The paper proposes that in order to close the research-practice gap, a multi-step empirical translation research framework is needed.

Keywords: Active learning; ICAP; Research-practice gap; How students learn

Correspondence should be sent to Michelene T.H. Chi, Mary Lou Fulton Teachers' College, Arizona State University, PO Box 37100, Mail Code: 3151, Phoenix, AZ 85069, USA. E-mail: mtchi@asu.edu

1. Introduction

2. A gap exists between research and practice

Thousands of studies have been carried out by scholars in the cognitive, education, and learning sciences relevant to how students learn. However, the findings from these studies have not trickled down to classroom implementations by practitioners and teachers, in both the K–12 grades and the postsecondary levels. This gap, referring to the lack of tangible impact of educational research knowledge in the classrooms (Schneider, 2018), has been recognized for decades among researchers, practitioners, and funders. Numerous types of accusations have been levied for why such a gap exists, such as

1. Research-based knowledge tends to be *too theoretical,* abstract, and general (McIntyre, 2005, p. 359) because the goal of research is "aimed at extending our understanding of the world around us and generating explanations, theories and predictions (Hirschkorn & Geelan, 2008, p. 2); whereas practical knowledge needed by teachers is context-specific."
2. Researchers address issues and questions *irrelevant* to practice and not important to teachers (Gutiérrez & Penuel 2014; Waxman, Freiberg, & Knight, 1986); that is, "it is researchers, not practitioner, who determine the agenda for educational research" (Hargreaves, 1996, p. 3).
3. Practitioners find the researchers' work *inaccessible* such as their writing style is full of jargons (Carnine, 1999); research findings are not communicated in easily digestible forms (Joyce & Cartwright, 2020, p. 1047), and not disseminated to practitioners (Osher & Snow, 1998, p. 256); nor disseminated in venues that practitioners frequent (Wilson & Corbett, 2000, p. 15).
4. Practitioners *lack time* to read up on the available research findings (Cox, Kahn, & French, 1985; Fleming, 1988); nor figure out "whether a program can work for them, and if so, what they need to put in place to get it to do so" (Joyce & Cartwright, 2020, p. 1046). In other words, the current situation makes it the job of the practitioners to find the appropriate research findings, resolve discrepancies among the research findings, and select the one appropriate for their specific context.

Although these accusations (*too theoretical, irrelevance, inaccessible,* and *lacking time*) are not false, they are levied at both the researchers and the practitioners, which led Wilson and Corbett (2000 p. 15) to ask: "Whose responsibility is it to translate research into practical implications?" It seems obvious that both researchers and practitioners must be involved in doing the translation work, although the researchers may have to shoulder a lion's share of the burden. The challenging question is how? What does it take to close the gap between research and practice?

3. Existing efforts to address the gap

The existence of a gap over the past several decades has prompted substantial progress in addressing the accusations through at least four major efforts, all primarily supported by the U.S. Institute of Education Sciences (IES). The first most impactful trend in addressing the *too theoretical* issue is that research has become more focused on evidence-based findings that work for improving learning outcomes in classrooms.

The second effort, addressing the *irrelevance* issue, is to form researcher-practitioner partnerships, which are long-term collaboration that involves multiple studies and projects. In this approach, researchers and practitioners negotiate problems of practice, such as what curriculum to use; and they also collaboratively co-design and test solutions (Coburn & Penuel, 2016). Thus, this approach clearly has the advantage of addressing critical problems that teachers face, thus avoiding the *irrelevance* issue.

IES has supported research-practitioner partnership grants for close to a decade, at $40 million per year (Schneider, 2018, IES Blog http://ies.ed.gov/blogs/) and continues to support the program without new grants competitions. See 2020 blog post from IES Director: https://ies.ed.gov/director/remarks/2-4-2020.asp. But it is not clear how to evaluate its success or determine what conditions might have led to that success.

The third effort, also supported by IES, is to create an organization, What Works Clearinghouse, that evaluates the available research, summarizes the results, and advertises interventions that have proven efficacious in randomized controlled studies that were carried out by the researchers. This would help in addressing the practitioner's *inaccessible* and *lack of time* issues. There are about a dozen topics covered in What Works Clearinghouse, including three in the major content domains (literacy, math, science) and three for the three age groups (early childhood, K–12 grades, and postsecondary). In addition, there are half a dozen other topics, such as children and youth with disabilities, English learners, behavior, path to graduation, charter schools, and so forth.

For each topic, there are many subtopics. In the case of literacy, there are over 100 subtopics, such as peer-assisted learning strategies, and within that subtopic is a list of five studies that meet rigorous design standards. Then for these five studies that meet design standards, an index of its effectiveness is provided, or sometimes no significant improvements are found. Thus, this is a huge database of successful and unsuccessful studies, and it seems to be a pretty daunting task for a practitioner to figure out which intervention she should try to implement.

Effective results approved by What Works Clearinghouse so far primarily are ones focusing on implementation of new classroom curriculums, professional development programs for teachers, school-level changes such as charter schools with unique schedules or high teacher expectations, and supplemental technology programs for use in the classroom. But most notably, what needs to be pointed out here is that in searching this huge database, we are hard-pressed to see the influence of the science of learning (or principles from the science of learning, to be explained in Tables 1 and 2) in these works. This merely confirms the lack of impact that the science of learning has had in educational practice.

TABLE 1
Seven Principles or Recommendations Taken from the Institute of Education Sciences Practice Guide

How to Design or Present Instructional Materials to Overcome or Enhance Limited Human Processing Capacity	How to Elicit, Affect or Train Students in a Skill that Requires Effortful Processing
(1) Space learning over time	(7) Use prompts and deep questions to elicit students' explanations
(2) Interleave worked example solutions with problem-solving exercises	
(3) Combine graphics with verbal descriptions	
(4) Connect and integrate abstract and concrete materials	
(5) Use quizzes to re-expose students to key concepts	
(6) Help students allocate study time efficiently (only with moderate rather than strong support)	

Note. From *Organizing Instruction and Study to Improve Student Learning* (NCER 2007-2004) by Pashler et al., 2007, Washington DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ncer.ed.gov.

TABLE 2
The 25 Principles with Substantial Empirical Support Classified into Two Groups

How to Design or Present Instructional Materials to Overcome or Enhance Limited Human Processing Capacity	How to Elicit, Affect or Train Students' in a Skill that Requires Effortful Processing
#1 Contiguity effects	#08 Organization effects
#2 Perceptual-motor grounding	#17 Explanation effects
#3 Dual code and multimedia effects	#18 Deep questions
#4 Testing effect	#22 Imperfect metacognition
#5 Spacing effect	#24 Self-regulated learning
#6 Exam expectations	
#7 Generation effect	
#9 Coherence effect	
#10 Stories and example cases	
#11 Multiple examples	
#12 Feedback effects	
#13 Negative suggestion effects	
#14 Desirable difficulties	
#15 Manageable cognitive load	
#16 Segmentation principle	
#19 Cognitive disequilibrium	
#20 Cognitive flexibility	
#21 Goldilocks principle	
#23 Discovery learning	
#25 Anchored learning (design challenging practical problem)	

Note. From *25 Learning Principles to Guide Pedagogy and the Design of Learning Environments* by Halpern et al., 2007, Washington, DC: Association for Psychological Science.

The fourth effort to close the research-practice gap is to identify which phenomena already uncovered in decades of laboratory studies in the psychology and education literature are in fact robust and derive their instructional implications. This approach consolidates and synthesizes evidence-based findings, followed by recommendations for how to implement them. This (referred to here as the) "consolidating" of empirical findings approach not only addresses three of the four accusations named above (all but the *irrelevance* issue), but this approach is extremely appealing because it is evidence-based. It will be expanded below.

3.1. The "consolidating" approach

The consolidating approach distills and synthesizes laboratory findings to identify robust phenomena relevant to education. An early attempt using this approach was commissioned by the IES for a panel of seven experts, charged to identify and describe robust findings in a way that will communicate evidence-based advice to practitioners in *actionable* recommendations (see Pashler et al., 2007). The panel arrived at a list of seven robust findings (shown in Table 1) and formulated actionable recommendations, each recommendation described in two or more pages of text with explicit advice for what to do. A similar second consolidation effort was made around the same time by Halpern, Graesser, and Hakel (2007). They assembled a collection of 25 robust principles to guide pedagogy and the design of the learning environment shown in Table 2.

More recently, a third consolidation effort was undertaken by a group of leaders of colleges of education called Deans for Impact (2015), composed of leaders of colleges of education, a national non-profit organization committed to data-driven improvement with empirical validation of effectiveness to transform the field of education preparation. In collaboration with a cognitive scientist (Dan Willingham at the University of Virginia) and a former middle school science teacher (Paul Bruno), they identified six key questions that all educators should be able to connect to their practical implications (Deans for Impact, 2015). Many principles do overlap with the prior two consolidating efforts.

Most recently, Kosslyn (2021) published a book, *Active Learning Online: Five Principles that Make Online Courses Come Alive* that also harvested "decades of knowledge about how students learn and applying this knowledge to remove learning." (p. 2). Thus, this consolidating approach seems to be the logical and popular route to take to close the research-practice gap.

3.2. Limitations of the consolidating approach

Although the consolidation effort to come up with a set of robust findings itself is commendable, this effort falls short on several fronts for the purpose of instruction. Each of these limitations is described below, using primarily the 25 principles listed in Table 2 as examples because there are many of them.

3.2.1. Skewed focus on instructional design and presentation of materials

Although not previously grouped, it might be insightful to categorize the seven or 25 robust findings into two topics as shown in Tables 1 and 2. Perhaps not mutually exclusive, but the principles listed on the left column of Tables 1 and 2 seem to be distinctly different from the ones on the right column. The principles in the left column are recommendations on how to redesign instructional materials or their presentations. Their goals seem to be to cope with the non-malleable aspects of human information processing capacities, by either *working around* the limitations of human information processing or *taking advantage* of the properties of human information processing capacity. For example, contiguity effects (#1, left column Table 2) *takes advantage* of the fact that it is easier for humans to recall related information if they are presented in proximity (so that they can be consolidated within memory capacity at the same time), perceptual-motor grounding (#2, Table 2) takes advantage of humans' ability to benefit from visualizing and manipulating concrete images or objects, and dual code and multimedia effects (#3) takes advantages of humans' ability to benefit from the information presented in redundant sensory (visual and auditory) or verbal and pictorial modalities. The principles on the left column also *work around* limitations of the human information processing capacity, such as that multiple testing will prevent students from forgetting (testing effect, #4), spaced instruction over time is also likely to prevent forgetting (spacing effect, #5), and presenting and segmenting information also overcome memory overload (managing cognitive load, #15; segmentation principle #16). Another way to characterize these principles is to say that they either work around or support the non-malleable aspects of human information processing so that working around their limitations or supporting their processing properties should benefit all learners.

These modifications of instructional materials or their presentations seem to allow students to process information without conscious effort. Even for recommendations such as interleave worked examples with problem-solving (#2, Table 1), designing and presenting such interleaved problems naturally lead students to solve the problems by analogy to the worked examples (i.e., by mapping the to-be-solved problem statements to the worked-out example statements and then copying the solution steps). The same is true for multiple examples (#11, Table 2) because presenting multiple examples inadvertently allows students to notice the varying conditions within the multiple examples under which an equation or a principle applies.

The recommendations on the right columns of Tables 1 and 2, on the other hand, are subtly different in that, even though they may require a redesign of instruction, and they seem to involve either training or eliciting from students the application of a strategy or a learning skill. These skills seem more effortful for students to undertake. For example, for the three recommendations (organizing #8, explaining #17, and asking deep questions #18), students have to explicitly and effortfully synthesize the materials, generate explanations, and ask deep questions. Imperfect metacognition (#22) and self-regulated learning (#24) also both require training and conscious efforts to undertake them. Thus, the five recommendations listed on the right column of Table 2 seem quite different from the ones on the left column in that students are either taught, trained, or elicited to use a skill.

Although both sets of recommendations seem to benefit students, this distinction leads to two very different cost and consequence. The cost is that it takes substantial efforts and know-how on the part of the instructors to figure out how to redesign and restructure instruction or instructional materials. Although some instructional redesigns are easy to implement, such as testing effect and the spacing effect (#4 & #5, Table 2), others are quite difficult. For example, to create materials that can cause cognitive disequilibrium (#19), not only do instructors have to first assess what prior false beliefs students have on a specific concept, but then they have to figure out how to modify instructional materials to cause conflicts; moreover, different confrontation materials may have to be designed to tailor to different sets of students (i.e., those with similar prior beliefs). Once conflicts are created, it is even more challenging to figure out how instructors should resolve the conflicts. This is quite a daunting task that researchers have not even figured out. Similar assessment and redesign efforts are needed also for the Goldilock principle (#21) and cognitive flexibility (#20). The consequence of this cost is that once redesigned, the modified instructional materials are mainly applicable only to that one concept or topic. For example, the materials needed to cause conflicts or disequilibrium on one concept (such as diffusion of oxygen and carbon dioxide), will not likely be the same as the materials needed to cause conflicts about how the phases of the moon work. This cost of the complexity of redesign of instructional materials coupled with the consequence that the redesigned instructional materials are not generalizable suggests that the recommendations listed in the left column, with a heavy emphasis on the instructional redesign, will require a substantial amount of training before practitioners can apply them; moreover, the redesigned materials may have limited generality.

In contrast, although some design and training efforts may be required for the principles on the right column, at least they are presumably transferrable across content domains. That is because the principles on the right column essentially recommend training or eliciting students on a learning skill, and once learned, such a strategy or learning skill can, in principle, be adopted or applied across content domains. For example, once students are trained to ask deep questions, they may be able to ask deep questions in multiple classes.

In summary, there seems to be an important difference among the principles listed in the two columns in Table 1 and 2. The difference is that the recommendations on the left column (80% of them) put the burden on the instructors and require them to redesign their instructional materials and presentation, thereby requiring researchers to provide more thorough and careful specifications to make the recommendations actionable (see the next topic below). Thus, the consolidating approach inadvertently skewed the efforts to close the gap by recommending costly and difficult instructional redesign with limited applicability across concepts or topics.

3.2.2. Recommendations not in fact actionable

Despite attempts at making the recommendations *actionable*, they are in fact not as actionable as they seem. In the preceding section, we described how difficult it is in fact to redesign instructional materials based on the principles, such as designing materials to create cognitive conflicts (cognitive disequilibrium, #19). The difficulty of designing training for the principles on the right column may be equally difficult for instructors to implement, based on the advice given, to be illustrated for the principle of explanation effects (#7 in Table 1, #17 in Table 2). Explanation effects essentially stated that teachers should “help students build explanations by asking and answering deep questions.” The idea was for teachers to “identify deep-level questions that they can use to prompt students to reason about underlying explanatory principles." This recommendation included three explicit advices for teachers: (a) "encourage students to think aloud in speaking or writing their explanations"; (b) "ask questions that elicit explanations," and one way to ask such questions was to use question stems such as *why, what caused X, what if, what-if-not, how does X compare to Y*; and moreover (c) "ask questions that challenge students' prior beliefs and assumptions, thereby challenging the students to consider deeper explanatory mechanisms and principles."

Perhaps we can liberally assume that the first advice on how to support and encourage students' self-explanations may not be difficult to implement, given that there are maybe a hundred or more laboratory studies that have done so, and the prompts used can be generic or content-free so they can be readily designed (Chi, in press), but the second advice of identifying and generating deep-level questions is far less actionable (Graesser, Ozuru, & Sullins, 2010). Not only do teachers find it difficult to generate deep questions without further guidance (Chi et al., 2018), but there is an insufficient clear-cut operational definition of what constitutes a deep question and how to generate one, other than using question stems (Graesser & Person, 1994). The third advice, suggesting that teachers ask questions that challenge students' prior naïve beliefs, may be even more difficult to implement, in that it is not clear how teachers know all of the students' prior naïve beliefs (Chi, 2013), nor know how to design instruction that will confront and remediate prior naive beliefs. Challenging students' prior naïve or misconceived beliefs is an extremely thorny problem that decades of research have not resolved (Chi, under review). Thus, the description of this seventh recommendation seems woefully insufficient as "actionable" pieces of advice that teachers can implement.

3.2.3. Indefinite growth

The list of 20 principles on the design and presentation of the materials (left column of Table 2) can grow indefinitely with more studies emerging with robust findings but with varying conditions. For example, instead of manipulating the presentation of multiple worked-out solution examples (#11 in Table 2), there are now also studies that have presented incomplete examples (i.e., examples with missing steps, Atkinson, Renkl, & Merrill, 2003), incorrect examples (Booth, Lange, Koedinger, & Newton, 2013), and more studies on examples interleaved with problem-solving (Van Gog, Kester, & Paas, 2011), and so on. Thus, there could be an infinite number of variations of the conditions of learning as determined by the presentation mode, the format, the sequencing of the instructional materials, and so forth. As the list of 20 principles grows, it will be more and more difficult for practitioners to distill from the long list what they need to know, and how to choose which principle to use and when, even if they have the time to invest in exploring their options.

3.2.4. A Lack of deep coherence

Finally, perhaps the most serious fourth limitation of the consolidating approach concerns a deeper lack of coherence among these lists of principles, related to the *inaccessibility* issue. Although the literature talks about *inaccessibility* in terms of jargons, communication, and dissemination issues, here, we will raise a more serious concern of the randomness or lack of coherence of the arrived set of robust recommended findings, making them in effect inaccessible. Clearly, there is not an apparent relation among the principles listed in the left column.

For example, superficially, the testing effect (#4) seems to facilitate recall by allowing students to retake a test, thereby forging more robust memory; it may also motivate students to seek additional information on what they have missed. But is there a relationship between this testing effect principle (#4) and the desirable difficulties principle (#14)?

Similarly, for the principles in the right column (Table 2), some of them address how to elicit or affect how students process information by the actions they take (such as self-explaining #17, asking deep questions #18), but there is no coherent theory that explains why those student actions enhance student learning or whether they are equally effective. For example, is it better to ask students to explain (#17) or to space studying and testing (#5)? Which recommendation achieves deeper learning?

There is no coherence in interpretation even within a single recommended activity or robust phenomenon. This can be illustrated with the learning benefits of concept mapping, a skill not listed in the tables but does have replicated findings; thus, it can be conceived of as a robust phenomenon (Chi & Wylie, 2014). Concept mapping is the activity of mapping out on paper or whiteboard the information and relations of information that are being learned. Concept mapping usually consists of drawing concept nodes and then indicating the relations among the nodes. The relations can be simple links (e.g., is related to) or more complicated links (e.g., is a part of, is caused by, etc). Concept mapping can be implemented in several different ways: Sometimes the nodes are provided but not their placement on the paper nor the links; other times the links are also provided and the student needs to indicate only what nodes are related by those links.

Suppose an eager teacher wants to know how to implement concept mapping and found these seven published articles about it, as shown in Table 3 (they are randomly selected for this illustration), and is willing to dedicate her precious time to read them. Each of the studies cited in Table 3 manipulates a concept-mapping condition in a different way and compares it to a different control condition. For example, study 2 (in Table 3) compares *copying* a concept map, which is better for learning than *reading* a concept map (Willerman & Mac Harg, 1991); whereas study 6 manipulates *correcting* a concept map and compares it to *reading* a text (Chang, Sung, & Chen, 2002). Even though these seven studies clearly lead to the robust conclusion that concept mapping is better for learning than many alternative tasks (used in the control conditions), there are several challenges from an adoption and practical implementation perspective.

First, some studies can be adopted and implemented more readily, while others may require more changes. For example, study 1 (in Table 3) shows that concept mapping plus listening to a lecture is better for learning than listening to lectures only (Schmid & Telaro, 1990). This can be easily implemented, such as after lecturing, give students a concept-mapping task because this is an additional task. But in study 6, the manipulation is to have students correct a concept map, which becomes more complicated to adopt and implement because no clear guidelines have been provided on how best to make a concept map incorrect. For example, is it better to have two concepts linked incorrectly or to have two concepts linked when they should not be linked at all?

Second, in addition to not knowing exactly how to implement an intervention based on the findings of specific studies (since various manipulations seem to result in different benefits

TABLE 3
A Random Sample of Seven Studies Comparing a Concept-Mapping Condition with Different Variations, and its Effectiveness Compared to a Control Condition (> or = signs)

	Activity 1	*Interactive, Constructive, Active, Passive* (ICAP) Mode		Activity 2	ICAP Mode	Study Reference
1.	Concept mapping + lecture	C	>	Listening to lecture only	P	(Schmid & Telaro, 1990)
2.	Copying a concept map	A	>	Reading a concept map	P	(Willerman & Mac Harg, 1991)
3.	Collaboratively building maps	I	>	Individually building	C	(Okebukola & Jegede, 1988; Czerniak & Haney, 1998)
4.	Collaboratively building maps	I	=	Collaboratively building maps with two additional resources	I	(Van Boxtel et al., 2000)
5.	Concept mapping	C	>	Reading	P	(Guastello, Beasley, & Sinatra, 2000)
6.	Correcting a concept map	C	>	Reading text	P	(Chang et al., 2002)
7.	Concept mapping from generating nodes and links	C	>	Concept mapping from selecting nodes and link	A	(Yin et al., 2005)

as shown in the studies listed in Table 3,), for a practitioner to read all seven articles listed in Table 3 brings other challenges. Based on these seven findings, which prescriptive advice is better to adopt? The choice is challenging because teachers will not understand why “concept mapping by generating nodes and links” is better for learning than “concept mapping by selecting nodes and links” (#7, Yin et al., 2005). Nor will they be able to choose among “concept mapping from generating nodes and links” (#7 in Table 3,), or “collaboratively building concept maps” (#4, in Table 3, Van Boxtel, Van der Linden, & Kanselaar, 2000), or correcting concept maps individually (# 6 in Table 3, Chang et al., 2002) since they cannot judge the comparative advantages of various manipulations of concept mapping when compared across studies.

Third, what about studies with null effects as shown in #4 in Table 3 (Van Boxtel et al., 2000). Why is providing two more additional resources not helpful for learning? Should this result be considered a lack of restriction on the boundary condition of how many resources are helpful?

Fourth, a larger question for teachers might be: Should they ask students to concept-map or use another learning strategy altogether, such as self-explaining? That is, which learning strategy is better, or does it not matter? None of the studies in Table 3 allude to such a contrast with other learning strategies.

These problems in the literature (non-systematic comparisons of multiple variations of an intervention activity, discrepant results, null results, etc.) compound the teachers' *lack of time* issue, in that it is unrealistic to expect teachers to massage the literature and translate a phenomenon or intervention, as it is difficult to interpret and know which manipulation of the phenomenon to adopt, or how to modify a manipulation to suit their own context. In short, this suggests that what is sorely needed is a theory on how students learn that can provide a coherent interpretation for all these manipulations and their comparative advantages as well as discrepant and null results. Having a theory will help teachers understand and choose the appropriate interventions as well as redesign them tailored to their own needs.

In summary, although the consolidation efforts all succeeded in synthesizing robust findings, this approach has many shortcomings for the purpose of prescription for instruction, such as skewing recommendations toward instructional design to overcome or support non-malleable human processing capacities, prescription not as actionable as it appears, potential for indefinite growth, and most seriously, a lack of explanatory coherence among the principles. Thus, a research-practice gap continues to exist even after robust phenomena have been identified from decades of laboratory research.

4. An alternative approach: Develop and translate a specific theory

For the many reasons mentioned above, instead of identifying and consolidating robust findings into a discrete list of recommendations for classroom practice, our approach is to develop a theory of how students learn that can account for many robust phenomena. The assumption is that, having a parsimonious and simple theory about how students learn that practitioners can easily understand could potentially provide a coherent interpretation for an infinite number of robust phenomena that might have implications for instruction. Moreover, armed with an understanding of a theory of how students learn, practitioners might be able to (a) design their own interventions tailored to their own classroom needs, providing them with autonomy and flexibility); (b) generalize it to whatever content areas, grade level, and context they choose; and (c) feel ownership of their revised lesson plans with a sense of control and autonomy (without feeling like their school is imposing another new approach or strategy).

Our call for translating a theory for practitioners is different from a similar call by Willingham (2017), who suggested that practitioners need one theory, by that he meant needing one theoretical perspective or approach, either a behaviorist, information processing, or a constructivist perspective. Our call is the need to develop and translate a specific theory, in this case, of active learning. Below, we first describe our theory of active learning, then what it takes to translate it for practical applications.

4.1. The Interactive, Constructive, Active, Passive (ICAP) theory of active learning

Active learning refers generally to how students engage with instruction (or instructional materials). Before delving into describing ICAP in more detail, there are two suppositions within ICAP that distinguish it from other assumptions and definitions of active learning.

First, ICAP claims that active learning can be segregated into four types, as opposed to discriminating it in a binary way as either "active" (meaning engaged) or "passive" (not engaged, usually meaning just paying attention), in the literature at large (Freeman et al. 2014). When defined in a binary way, "active" learning means anything else students do that is more than paying attention only, such as solving a problem. Active learning and ICAP both exclude disengagement, such as sleeping, goofing off, or any ways that are not even paying attention. Thus, paying attention, in practitioners' minds, is the loftiest attribute of being engaged with instruction.

Second, ICAP claims that the four ways that students can engage with instruction can be detected by *overt behaviors*. Although behaviors have been dismissed by the cognitive revolution for five decades or so, the assumption here is that behaviors might be salvaged and adequate for discriminating four *modes* or types of cognitive engagement rather than using behaviors to reflect precise thinking processes, which behaviors cannot do. The use of overt behaviors as an indicator allows classroom teachers to detect visually how students are behaving, thereby engaging. Although this is a coarse index of students' engagement, if necessary, students' *Active* or *Constructive* engagement modes can be discriminated more precisely by comparing the information contained in students' outputs with the information contained in the instructional materials. The similarity or discrepancy between the information contained in the students' outputs and the information presented or contained in the original instructional materials can determine more accurately whether students are engaging in the *Active* mode or the *Constructive* mode, respectively.

In the next section, we first describe our hypothesis of the elementary processes involved in learning, followed by a description of ICAP's operational definitions of the four engagement modes, the ICAP hypothesis and its predictions, and culling support for the hypothesis from published evidence in the literature.

4.1.1. The elementary knowledge-change processes of learning

ICAP is a theory of how students learn in the context of instruction (e.g., teachers' explanations, feedback, activities) or instructional materials (textbooks, tests, assignments). Both will be referred to more broadly as instruction. ICAP assumes that learning can be conceived of as the occurrences of these six elementary processes that have been well-established in the psychology literature. Since learning is causing some changes in one's prior knowledge, these can be referred to as knowledge-change processes and they are *storing, activating, connecting, changing, inferring, and reflecting.* That is

1. Externally presented new information can be *stored* during the encoding processes (when external information is being taken in), either in isolation or connected with other prior knowledge.
2. External information can *activate* related prior knowledge in memory.
3. New information can be linked to or *connected* with prior knowledge, or prior knowledge units can be connected with each other.
4. Prior knowledge can be *changed* or revised.

5. New ideas can be *inferred* from: prior *stored, activated,* or *revised* knowledge, or from prior knowledge *connected* with new information, and so forth.
6. *Reflecting* can be conceived of as the process of reactivating some prior knowledge in order to examine it, and perhaps making further comparisons, connections, inferences, revisions, and so forth.

Knowledge change processes 3 to 6 essentially help a learner build a correct representation or understanding of instructional information. Because we cannot tell in practice which of the knowledge-change processes 3 to 6 are occurring exactly, the term *infer* is used below to broadly include *connect, revise, reflect,* and to relay the idea that new knowledge is produced. Thus, for a practical discussion of learning, the knowledge-change processes reduce to three basic ones: *store, activate,* and *infer.*

In any learning situation, any number of these processes may be active, either in isolation or in combination with the other processes. The following is a simple rendition of when these knowledge-change processes operate during learning. When new information is being presented, in order to encode and *store* new information, not in isolation but connected with relevant prior knowledge (so as not to store knowledge in isolation, leading to inert knowledge, Whitehead, 1929), relevant prior knowledge must first be *activated*. Once relevant prior knowledge is activated, it has to be examined or *reflected upon* to see how the new information fits in (e.g., which piece of prior knowledge should a new piece of information be *connected* with). If no piece of knowledge for connection seems suitable, or if the new information contradicts prior knowledge, further *reflection* might be needed to determine how prior knowledge can be *revised* before connecting it with new information. If a connection is successful, further *inferences* can also be generated. The point is that several of these knowledge-change processes can occur rapidly and continuously during learning, and it is impossible to capture precisely which processes are occurring in practice.

4.1.2. Overt behavioral indicators of four modes of engagement

Because teachers cannot tell, in an authentic learning environment (such as a classroom), which knowledge-change processes students are thinking, so how can this challenge be resolved in order for teachers to know how students are engaged? One way is to map what students *do* while engaged with instruction that is visible behaviorally to the invisible knowledge-change processes they are undertaking. Accordingly, a taxonomy was proposed that specified roughly four distinguishable types or modes of student overt behaviors (along with the concrete outputs they produce), and these four modes were hypothesized to map to some of the knowledge-change processes. Based on which knowledge-change processes might be operating for each mode of external behavior, four levels of learning outcomes were predicted. While recognizing that such mapping cannot be perfectly accurate 100% of the time, it might be good enough for an authentic classroom environment. The four modes of student behaviors reflecting four levels of thinking are capitalized and italicized and can be illustrated below using the instructional contexts of lecturing, showing a video online, or providing a text. What students do in the context of each instructional activity is italicized.

The lowest mode of learning is reflected in what we labeled as *Passive/Attending* behavior, consisting of students *listening* to a lecture, *watching* a video, or *reading* a text silently. This *Passive/Attending* mode is more likely to elicit direct encoding and *storing* of new information without necessarily connecting it with prior knowledge.

The second mode of learning is reflected in *Active/Manipulating* behavior, consisting of students doing something with the instructional materials, such as *underlining* some text sentences while reading, *copying* the examples from the board or PowerPoint that the teachers wrote. The key in *Manipulating* is that students are not producing any new information beyond what was already presented in the instructional materials (i.e., the underlined sentences were already in the text), but underlining them simply emphasized them for more focused attention. Thus, what this manipulation might do cognitively is to activate relevant prior knowledge relevant to the underlined sentences, which then allows the learner to selectively focus more attention on the manipulated (i.e., underlined) parts of the instructional materials.

Activating relevant prior knowledge will have three important advantages: First, activating it will strengthen the activated prior knowledge so that it will be retrieved more easily in subsequent situations. Second, activating it followed by connecting it with new information will make the prior knowledge about the topic more complete or more enriched. Third, activating it may allow the learner to recognize that it needs to be revised before it can be connected to new information. Thus, *Manipulating* behavior is considerably beneficial to learning, compared to simply *Attending* and not doing anything else. This explains why calls for hands-on instruction in K–12 and learning-by-doing in higher education remain popular (Reese, 2011) because they do increase learning, compared to paying attention only.

The third mode of learning is reflected *Constructive* behavior, consisting of students *Generating* some information beyond what was already presented in the instructional materials. For example, as you read, if the incoming information causes you to *ask* a question or *come up with* another example, or wonder whether *this means X*, you are generating new information not already presented. Essentially in generating, you are building on what was already presented by inferring new ideas, new connections, new properties of an entity, and so forth, using various common inferring processes. Generating inferences while learning is distinct from generating schema-based inferences during text comprehension as described in Chi (in press).

Sometimes visible engagement behavior is ambiguous with respect to whether it reflects *Manipulative* or *Generative* knowledge-change processes. For example, when a student takes notes, it can indicate either *Manipulating* or *Generating* because the behavior of *taking* notes looks the same overtly. However, the two levels of learning can be further discriminated by comparing the notes to the instructional information. Examining the notes can confirm whether they contain new information beyond what was presented, such as when students are *taking notes* in their own words (which is *Generative* or a *Constructive* mode) as opposed to *copying* notes (which is *Manipulative* or an *Active* mode). Thus, behavioral ambiguity can be resolved by examining the outputs students produce, on the rare occasions when teachers need an accurate resolution of such ambiguities.

The fourth mode of learning occurs in *Interactive* behavior, which means *Collaborating* with another peer. However, for optimally effective collaboration that can enhance learning and creativity, each partner must be *Generative* not only by extending beyond her own prior contributions but also *Generative* by extending beyond her partner's contributions, referred to as *co-generative* or *co-constructive* way of interacting. Thus, the benefit of collaboration for learning is not merely that there are two peers with two sets of prior knowledge or perspectives, but that each peer can build on or extend (by inferring) not only the other peer's knowledge but more importantly, the other peer's inferred knowledge, in a way that the other peer did not consider. This means peer1 generated an inference X that did not occur to peer2 to infer, but peer2 can generate a subsequent inference Y beyond X, which would not have been generated had peer1 not generated X initially. This subsequent building upon an inference that one did not initially infer suggests that collaborating could potentially be more novel and creative because the outcome or solution is not one that either partner could have created alone. Thus, collaborative learning has the potential to foster an understanding that goes beyond what a person could achieve working alone, but only if the collaboration entails a co-generative type of interactions.

There are five important unique aspects to our theoretical approach. The first unique aspect is that our definition of active learning is based on the two criteria of students' visible behaviors and concrete outputs. This overcomes the constraint that teachers cannot see what students are thinking as thinking processes are invisible. The second unique aspect is that the implications of ICAP are applicable to all students, all grade levels, and for all content domains. For example, all students can be *Constructive/Generative*, regardless of their background knowledge. This is because the role of being *Constructive* is to make sense of the new information for oneself by adding, connecting, revising, essentially building one's own representation or understanding of this new information. Thus, everyone can benefit from this building activity, regardless of the status of one's initial representation. The third benefit of defining students' active learning mode on the basis of their outputs is that this operational definition can guide teachers to evaluate their own design of student activities, to determine what mode of student outputs their designed activities elicit (i.e., they can just compare the anticipated student response with the presented instructional materials). For example, ICAP's easily defined rubric, requiring only a comparison of what is generated to what is presented instructionally, can teach teachers how to design a deep question: A deeper question can simply be one that elicits generative responses from students, containing information that has not been presented instructionally. Recall earlier we stated that teachers had difficulty understanding and generating "deep" questions, and the research literature has not come up with a concrete definition. However, ICAP's definition of a *generative* question can be an adequate substitute for a "deep" question, since generative questions tend to be deeper questions. Fourth, ICAP takes a student-centered approach, in contrast to a teacher-centered approach. A student-centered approach means we consider what the students do to optimize learning rather than what teachers do to optimize learning, such as give feedback, space learning over time, or redesign instructional materials in ways recommended by the principles in the left column (of Tables 1 and 2). That is, we prescribe what teachers should do in instruction based on what ICAP mode a teacher's activity elicits in students' engagement. The most important fifth unique aspect of ICAP is that it assumes that teachers can be taught to understand ICAP, and based on their understanding, they can redesign their own lesson plans and in multiple topics. Thus, ICAP is a theory that employs unique suppositions.

4.1.3. The ICAP hypothesis

Mapping these four types of overt learning behaviors to the six plausible-knowledge-change processes shown above allows us to predict that learning decreases in this I>C>A>P direction. That is, ICAP stands for and predicts that *Interactive* or *Collaborative* learning behavior is better for learning than *Constructive* or *Generative* behavior carried out solo. *Generative* behavior is superior for learning than *Active* or *Manipulative* behavior, which in turn is superior to *Passive* or *Attentive* behavior. Thus, based on distinct externalized student behaviors, teachers can tell more or less how actively students are processing the learning materials. If necessary, for ambiguous cases, the classification of overt behaviors can be further confirmed by examining students' outputs. The implication of ICAP is that pairwise comparisons should therefore hold by transitive inference. For example, ICAP implies that C > P, I > A, and so forth.

There is one very important caveat with respect to the *Interactive* mode, which is that in order for the *Interactive* mode to be more beneficial for learning than the *Constructive* mode, it must have both partners not only each be *Generative,* but they must build on top of the partner's contributions so that they are both *co-generative* as explained above. That is, the definition of *Interactive* in ICAP, as noted above, is *Collaborative* in the *co-generative* sense. In reality, however, the majority of collaborative learning interactions are not *co-generative;* instead, one of these four non-*co-generative* patterns of interactions occur: (a) one peer dominates and is *Generative* while the other one is *Attentive*, (b) both peers are cooperating (meaning that they divide and conquer rather than mutually co-generate), (c) both peers work in parallel and are not interactive (i.e., each works alone and is *Generative*), or (d) they work in parallel but both are only *Manipulative.*

4.1.4. Evidence supporting ICAP's predictions

Once a hypothesis is generated, the theory's predictions must be tested. In the literature, this support is typically obtained by new studies undertaken by the researchers who created the theory. However, it is more powerful if a theory can be supported by studies in other laboratories. In the case of the ICAP theory, support was obtained from hundreds of studies carried out by others with only one study carried out in our laboratory. The studies in the literature were re-interpreted by mapping both the intervention condition and the control condition of a study to ICAP modes. Based on such mappings, we can then see whether the learning or performance outcomes manifest in the predicted *I>C>A>P* direction. Given that the majority of studies in the literature compare only two conditions, we can compare the six permutations of ICAP comparisons, such as whether *I>C, I>A, I>P, C>A, C>P,* and *A>P.*

There are three ways to show how ICAP can predict and re-interpret existing findings in the literature, whether they are null results, discrepant results, or non-comparative results. The first is in a broad stroke way by examining whether the results of a study are predicted by ICAP if the study's conditions can be cleanly mapped to an ICAP mode. Such re-interpretations can occur for classroom studies, laboratory studies with adults, developmental studies with young children, and even studies with toddlers. For a classroom example of such mapping, in a study by Henderson (2019), college students used clickers to respond to conceptual physics questions. After responding, they were either given an opportunity to verbally discuss their clicker votes with each other (mapped to the *Interactive* mode) or they received a supplemental lecture (mapped to the *Passive* mode) between clicker votes. Students who collaborated performed significantly better on the force concept inventory than students who listened to a lecture, consistent with the $I > P$ prediction.

ICAP can also explain the benefit of the classic testing effect (#4, Table 2). That is, retesting, which causes students to *retrieve* prior learned knowledge prompted by the test questions, is an *Active* activity in which students *Manipulate* by retrieving already learned information. Retesting is typically contrasted with presenting students with the same information twice. Presenting information allows students to be merely *Passive/Attentive;* thus, it makes sense that retesting is better for learning than representing information.

For an example of a study with young children, Legare and Lombrozo (2014) showed that when five-year-olds are shown a crank turning gears, the group that was asked to *explain* how that might happen, without getting any feedback (the explain group), performed significantly better in coming up with the correct causal explanation at the end than the group who was only asked to *observe* without explaining. This makes sense in ICAP because the explain group is being *Constructive/Generative*, whereas the observe group is merely *Passive/Attentive.* However, not all explainers actually attempted to *explain* the causal relation between the crank and the gears; some of the explainers only *described* the crank and the gears. For the describers, they performed significantly worse at the final explanation than the explainers because according to ICAP, describers were only *Active/Manipulative* because they provided no new information in their descriptions, whereas the explainers were *Constructive/Generative.*

ICAP can also explain findings in non-school and non-intentional learning contexts, such as the classic finding that a toddler's vocabulary growth depends on whether parents read to a toddler in a way that engages the toddler in talking about the story (which is *Constructive)* or not (the *Passive* mode, De Temple & Snow, 2003). It might not be too far-fetched to suggest an ICAP interpretation even for the context of eating at a dinner table. That is, studies have shown that that the vocabulary and intellectual development advantages observed in children who enjoy regular family meals are due to the *Constructive* nature of mealtime conversations (Snow & Beals, 2006). In short, such re-interpretation analyses of the findings in the literature have been done for hundreds of studies, many are cited in the two prior ICAP papers (Chi, 2009; Chi & Wylie, 2014), mostly for pairwised comparisons, with a couple of three-way comparisons, and our own four-way comparison (Menekse, Stump, Krause, & Chi, 2013). In short, ICAP provides a single unifying explanation for hundreds of studies in the literature.

A second way to validate ICAP's predictions is to show how ICAP can explain the discrepant, null, and non-comparative findings on a single phenomenon, such as concept mapping. Going back to Table 3, the two ICAP mode columns indicate the ICAP mode we have assigned to each condition, based on what students were asked to do in those studies. For example, *correcting* a concept map (Table 3, item 6) is a *Constructive* activity because students were asked to generate corrections for the concept map, so they were producing something not already presented in the learning materials. On the other hand, *copying* a concept map (item 2, Table 3) or *selecting* nodes and links (item 7, Table 3) is *Active* because no new information was provided when students copied and selected what was already presented. In looking at Table 3, the assigned mode predicts and confirms the direction of each study's findings. For example, in item 3, the prediction that collaboratively building maps facilitates learning more than individually building maps ($I > C$) is supported by findings from two studies (Czerniak & Haney, 1998; Okebukola & Jegede, 1988). Table 3 also shows that ICAP can explain non-intuitive findings, discrepant results, null results (e.g., item 4, in which equivalent results were found since both conditions were *Interactive)*, and predict comparisons that have not been studied yet (see Chi, 2009; Chi & Wylie, 2014).

A third way is to use ICAP to provide coherence in cataloging many distinct student activities. For example, we can now easily interpret that many of the robust findings listed in the right column of Table 2 are effective for learning because they are all *Constructive/Generative* activities, such as explaining, asking deep questions, discovery learning, and so forth. Moreover, ICAP can even make predictions about the relative effectiveness of the recommendations listed in the left column of Table 2. For example, the testing effects (#4) essentially re-exposes students to the same questions and asks students to *re-retrieve* information that they have stored and learned. Retrieval engages students in an *Active* mode since no new information is produced. This testing effect is effective for recall but probably not as effective for deeper learning as desirable difficulties (#14), since desirable difficulties require students to *reorganize* the presented information. Of course, the learning benefit depends on the mode of reorganization needed: If the reorganization merely requires students to *r-order* some sequencing of information, then that is engaging in the *Active/Manipulative* mode since no new knowledge is provided, but if the reorganization requires students to *integrate* two sources of information, then that is engaging in the *Constructive/Generative* mode since integrating produces new integrated knowledge. Thus, using ICAP's operational definition of whether some new information is produced, ICAP can in principle discriminate which engaging activity is better for learning, and make comparative predictions between recommended activities (such as *re-retrieving* #4 versus *synthesizing* #14) as well as comparative predictions within a recommended activity (such as *reordering* or *synthesizing*, for #14). Note that ICAP categorizes the mode of student engagement based on what the learners are asked to do as shown in the italicized verbs in this paragraph. Using ICAP to analyze the principles in Tables 1 and 2 further reinforces the concern that the consolidated principles skew the emphasis on designing instruction. This emphasis on instructional design overlooks the fact that some of the recommendations might involve only the *Active* mode, and K–12 teachers are already quite facile at designing *Active* activities. In fact, the majority of K–12 student activities are designed for the *Active* mode (Chi et al., 2018).

However, there are a few caveats to ICAP's predictions. One caveat is that being generative is not beneficial for a few content domains that have arbitrary rules, such as learning rules of grammar since there are no inferences or justifications that one can generate. The second caveat is that ICAP's predictions are restricted to those between ICAP modes but not within an ICAP mode because many other factors may affect the outcomes of comparing within a mode, such as the differences in the affordances or cognitive demands of two tasks within the same mode (Chi, in press). For example, suppose we are considering two *Constructive* student tasks: *asking questions* or *forming a hypothesis*. Asking a question seems to be cognitively less demanding than forming a hypothesis, which may require a great deal of consolidation of what was read and learned. One way to confirm that these two generative activities differ in terms of cognitive demands might be the use of time as a metric, such that forming a hypothesis most likely requires more time than asking a question so that forming a hypothesis may engender more learning than asking a question. ICAP cannot predict this difference in outcome because it requires additional analyses of the thinking processes beyond analyses of students' overt behaviors and outputs. However, since instructors only need to discern how engaged their students are in an authentic context, the coarse-grained discrimination that ICAP provides is adequate for that purpose.

4.2. How to translate ICAP, a theory of learning into a theory for instruction

Once a theory has been developed or a robust phenomenon identified (referred to as "the intervention"), in order to translate a theory or any robust phenomenon into practice, our claim is that at least six or seven additional research steps are involved. In a nutshell, these additional steps involve (a) having the researchers develop explanations of the theory or phenomenon to practitioners; (b) making sure that practitioners understand the theory or the phenomenon by developing an objective assessment of that understanding; (c) practitioners then further exhibit their understanding by revising and redesigning their instruction according to the theory or the phenomenon, with the researchers developing ways to assess the practitioners' redesigns for their accuracies; (d) then the practitioners implement their revised instructional lessons, and their implementation has to be evaluated for its fidelity, again often requiring creative new method and rubric of coding implementation fidelity; (e) then whether students comply and enact the activity as requested need to be assessed; (f) along with whether students then in fact learned significantly more, with proper pre-posttest assessments, and the final big step (g) is to see if the intervention would operate at scale, in multiple contexts with multiple student populations.

The steps laid out above constitute an empirical translation research framework that dictates at minimum carrying out these seven systematic and effortful steps in order for researchers and practitioners to translate a robust research phenomenon or a theory into actionable practice. Many challenges can arise within each step, revealing problems in situ that have not been previously addressed in research, offering exciting opportunities for researchers to come up with novel methods for analyses. Here are two examples. First, in order to evaluate and discriminate the ICAP mode of teachers' revised and redesigned activities or assignment, we had to devise a new coding rubric based on the verbs they used in the assignment to describe the activities. For example, for an assignment in mathematics of finding equivalent fractions, if a problem asked students to *match* from among the four options in the equivalent representation, then this is an *Active/Manipulative* activity as the choices are all given. However, if the assignment instruction said to *come up with* an equivalent fraction representation, then that is a *Constructive/Generative* activity. We discovered that the hundreds or so verbs teachers used could easily be classified as either *Active* or *Constructive,* with some *Passive* ones (Chi et al., 2018). Second, we can easily code teachers' questions as either *Active* (if they ask questions such as "What is the name of the city ….," as that information was already provided), or *Constructive* (such as "Think of another example…" as that new example was not already provided. These coding rubrics can then be converted to instruction to explain to teachers how to design problem assignments and ask deeper questions (Morris & Chi, 2020). Similarly, within this empirical translation research framework, teachers also need to be creative in coming up with adaptations if students are not complying. This level of commitment in research effort is needed in order to close the gap between research and practice.

A more detailed description of the translation steps as well as the example of translating ICAP into practice is published in *Cognitive Science* (Chi et al., 2018). Our first description of this theory with supporting evidence is published in *Topics in Cognitive Science* (Chi, 2009), and a subsequent follow-up paper is published in *Educational Psychologist* (Chi & Wylie, 2014). The Menekse et al. (2013) paper is our empirical test of the ICAP predictions across four modes. These four publications are companion pieces to the current paper. An initial version of around 300 PowerPoint slides has been created that explained ICAP for teachers. A sample mini-version can be seen on the author's website or on this link: https://education.asu.edu/sites/default/files/chilab/icap-how-students-engage-to-learn/index.html#/

5. Conclusion

That research findings do not trickle down into classroom practice is a gap that has been recognized by researchers, educators, and funders. Closing this gap has been addressed in four prongs. The first prong consists of more funding support for research that takes place in authentic classrooms. The second prong is to provide funding for research-practitioner partnerships. The third prong is to provide a database of successful classroom research, the What Works Clearinghouse. This clearinghouse includes a broad spectrum of work but mostly successful curricula and other instructional resources. The fourth prong is to take advantage of the decades and decades of research on human learning and teaching and distill ones that are robust and translate them into practical advices for practitioners. Referred to in this paper as the "consolidating" approach, this approach has the most appeal because it is based on evidence provided by solid, replicated findings. However, a closer analysis of this consolidating approach reveals that it has many limitations and comes up far short of implementable practices that instructors can actually use and apply. The limitations include: (a) an over-emphasis on an instructional redesign, thereby putting burdens on practitioners without providing them with precise actionable implementation procedures; (b) a lack of explanations for why certain intervention might work and others might not, the conditions under which they might work, and what their comparative advantages are; and (c) a lack of coherence overall.

To address the failure of the consolidating approach, it is suggested that translating a specific theory approach may overcome many of its shortcomings. The ICAP theory of active learning posits that there are four modes of active learning, definable by students' overt behaviors and outputs, and these four modes of engaging achieve different levels of learning, with the *Interactive* and the *Constructive* modes garnering deeper learning than the *Active* and *Passive* modes. Evidence supporting the predictions of the I>C>A>P hypotheses was alluded to and a few studies were briefly described to illustrate that ICAP can provide a unifying explanation for a range of studies across ages, content, and contexts. The point is that translating this theory into instruction for practitioners might be a more powerful and generalizable way of advising them because with an understanding of ICAP practitioners can design their own improvements to their lesson plans and apply improvements to various topics and content domains. This approach would empower teachers and instructors, and give them the autonomy and leeway to apply the theory's prescriptions themselves. Furthermore, teachers and instructors can also use ICAP to evaluate students' engagement behaviors and outputs, to assess their own pedagogy, to assess the utilities of technology tools, and so forth. This empowerment and autonomy are more likely to potentiate success in adoption and application by practitioners.

Translating a theory into practice requires implementing the multiple steps of the empirical translation research framework described above, involving researcher-teacher partnerships to systematically evaluate the adequacy of the researchers' explanations of a theory or phenomenon, explores the feasibility and fidelity of teachers' implementation, and assess students' compliance and learning outcomes, with final scale-up issues that have to be tested and resolved.

In short, the analysis presented in this paper with respect to what research is needed to close the gap between research and practice suggests that activities such as asking researchers to make their work more accessible in terms of communication clarity, to present at practitioners' conferences, and to promote more dialogs between researchers and practitioners, are harmless but superficial solutions for closing the gap. The goal of this paper is to advance the notion that in order to close the gap, we must do translation *research* and not simply state summarily at the end of a research report that this work has implications for education.

Acknowledgments

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grants R305A110090 and R305A150432 to Arizona State University. The opinions expressed are those of the author and do not represent the views of the Institute or the U.S. Department of Education. The author is grateful for comments provided by Erin Higgins, Program Officer, Institute of Education Sciences, as well as two anonymous reviewers and the editor Wayne Gray.

References

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, *95*(4), 774.

Booth, J. L., Lange, K. E., Koedinger, K. R., & Newton, K. J. (2013). Using example problems to improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction*, *25*, 24–34.

Carnine, D. (1999). Campaigns for moving research into practice. *Remedial and Special Education*, *20*(1), 2–35.

Chang, K. E., Sung, Y. T., & Chen, I. D. (2002). The effect of concept mapping to enhance text comprehension and summarization. *The Journal of Experimental Education*, *71*(1), 5–23.

Chi, M. T. H. (in press). The self-explanation principle in multimedia learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed.) Cambridge: Cambridge University Press.

Chi, M. T. H. (under review). Individualistic versus collective causality: A unifying framework for analyzing process-concepts and explaining Misconceptions. *Cognitive Science*.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. https://doi.org/10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H. (2013). Two kinds and four sub-types of misconceived knowledge, ways to change it and the learning outcomes. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (2nd ed., pp. 49–70). New York, NY: Routledge Press.

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, *42*, 1–56. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*, 219–243. https://doi.org/10.1080/00461520.2014.965823

Cox, P. L., Kahn, K. A., & French, L. C. (1985). *Making the match for use of educational information: Vol. 3. A study of clients of information service providers and their use of ERIC-based resources and services*. Andover, MA: The Network.

Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher*, *45*(1), 48–54.

Czerniak, C. M., & Haney, J. J. (1998). The effect of collaborative concept mapping on elementary preservice teachers' anxiety, efficacy, and achievement in physical science. *Journal of Science Teacher Education*, *9*(4), 303–320.

De Temple, J., & Snow, C. E. (2003). Learning words from books. In A. Van Kleeck, S. A. Stahl & E. D. Bauer (Eds.), *On reading books to children: Parents and teachers* (pp. 16–36). Mahwah, NJ: Lawrence Erlbaum Associates.

Deans for Impact (2015). *The science of learning*. Austin, TX: Deans for Impact.

Fleming, D. S. (1988, April). *The literature on teacher utilization of research: Implications for the school reform movement* [Paper presentation]. The Annual Meeting of the American Educational Research Association: New Orleans, LA.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*, 8410–8415.

Graesser, A., Ozuru, Y., & Sullins, J. (2010). *What is a good question?* In M. G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 112–141). New York, NY: Guilford Press.

Graesser, A. C., & Person, N. K. (1994). Question-asking during tutoring. *American Educational Research Journal*, *31*, 104–137.

Guastello, E. F., Beasley, T. M., & Sinatra, R. C. (2000). Concept mapping effects on science content comprehension of low-achieving inner-city seventh graders. *Remedial and Special Education*, *21*(6), 356–364.

Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to practice as a criterion for rigor. *Educational Researcher*, *43*(1), 19–23.

Halpern, D. F., Graesser, A., & Hakel, M. (2007). *25 learning principles to guide pedagogy and the design of learning environments*. Washington, DC: Association of Psychological Science Taskforce on Lifelong Learning at Work and at Home. Retrieved from http://www.adesignfor.education/wp-content/uploads/2019/02/25-lifelong-learning-principles.pdf

Hargreaves, A. (1996). Transforming knowledge: Blurring the boundaries between research, policy, and practice. *Educational Evaluation and Policy Analysis*, *18*(2), 105–122.

Henderson, J. B. (2019). Beyond "active learning": How the ICAP framework permits more acute examination of the popular peer instruction pedagogy. *Harvard Educational Review*, *89*(4), 611–634. https://doi.org/10.17763/1943-5045-89.4.611

Hirschkorn, M., & Geelan, D. (2008). Bridging the research-practice gap: Research translation and/or research transformation. *Alberta Journal of Educational Research*, *54*(1), 1–13.

Joyce, K. E., & Cartwright, N. (2020). Bridging the gap between research and practice: Predicting what will work locally. *American Educational Research Journal*, *57*(3), 1045–1082.

Kosslyn, S. M. (2021). *Active learning online: Five principles that make online courses come alive*. Boston, MA: Alinea Learning.

McIntyre, D. (2005). Bridging the gap between research and practice. *Cambridge Journal of Education*, *35*(3), 357–382. https://doi.org/10.1080/03057640500319065

Menekse, M., Stump, G., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, *102*, 346–374. https://doi.org/10.1002/jee.20021

Morris, J., &, Chi, M. T. (2020). Improving teacher questioning in science using ICAP theory. *The Journal of Educational Research*, *113*(1), 1–12.

Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology*, *126*, 198–212.

Okebukola, P. A., & Jegede, O. J. (1988). Cognitive preference and learning mode as determinants of meaningful learning through concept mapping. *Science Education*, *72*, 489–500.

Osher, D., & Snow, S. (1998). *Using technology to link families and schools to research-based information* [Paper presentation]. The Families, Technology, and Education Conference, Chicago, IL.

Pashler, H., Bain, P., Bottage, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (NCER 2007-2004). Washington DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ncer.ed.gov

Reese, H. W. (2011). The learning-by-doing principle. *Behavioral Development Bulletin*, *17*(1), 1–19. https://doi.org/10.1037/h0100597

Schmid, R. F., & Telaro, G. (1990). Concept mapping as an instructional strategy for high school biology. *Journal of Educational Research*, *84*, 78–85.

Schneider, M. (2018). Message from IES Director: How to make education research relevant to teachers [News blurb]. *Institute of Educational Sciences.* Retrieved from https://ies.ed.gov/director/remarks/11-14-2018.asp

Snow, C. E., & Beals, D. E. (2006). Mealtime talk that supports literacy development. *New Directions for Child and Adolescent Development*, *2006*(111), 51–66.

Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction*, *10*(4), 311–330.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology*, *36*, 212–218.

Waxman, H., Freiberg, J., & Knight, S. (1986, February). *Using research knowledge to improve teacher education: Teachers' perceptions of the value of educational research* [Paper presentation]. The Annual Meeting of the Association of Teacher Educators, Atlanta, GA.

Whitehead, A. N. (1929). *The aims of education*. New York, NY: Macmillan

Willerman, M., & Mac Harg, R. A. (1991). The concept map as an advance organizer. *Journal of Research in Science Teaching*, *28*(8), 705–711.

Willingham, D. T. (2017). A mental model of the learner: Teaching the basic science of educational psychology to future teachers. *Mind, Brain, and Education*, *11*(4), 166–175.

Wilson, B., & Corbett, D. (2000, February). *The prospects for disseminating research to a hungry field.* Report to the National Center for the Study of Adult Learning and Literacy, Boston, MA.

Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C., & Shavelson, R. J. (2005). Comparison of two concept-mapping techniques: Implications for scoring, interpretation, and use. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *42*(2), 166–184.