The different types of contributions to knowledge (in CER): All needed, but not all recognised

STEVE DRAPER, School of Psychology, University of Glasgow, United Kingdom
JOSEPH MAGUIRE, School of Computing Science, University of Glasgow, United Kingdom

The overall aim of this paper is to stimulate discussion about the activities within CER, and to develop a more thoughtful and explicit perspective on the different types of research activity within CER, and their relationships with each other. While theories may be the most valuable outputs of research to those wishing to apply them, for researchers themselves there are other kinds of contribution important to progress in the field. This is what relates it to the immediate subject of this special journal issue on theory in CER. We adopt as our criterion for value "contribution to knowledge".

This paper's main contributions are:

- A set of 12 categories of contribution which together indicate the extent of this terrain of contributions to research.
- Leading into that is a collection of ideas and misconceptions which are drawn on in defining and motivating "ground rules", which are hints and guidance on the need for various often neglected categories. These are also helpful in justifying some additional categories which make the set as a whole more useful in combination.

These are followed by some suggested uses for the categories, and a discussion assessing how the success of the paper might be judged.

CCS Concepts: • **Social and professional topics → Computing education**.

Additional Key Words and Phrases: theory, computing education research, taxonomy, contributions to knowledge

ACM Reference Format:
Steve Draper and Joseph Maguire. 2021. The different types of contributions to knowledge (in CER): All needed, but not all recognised. *ACM Trans. Comput. Educ.* 1, 1, Article 1 (January 2021), 35 pages. https://doi.org/10.1145/3487053

1 INTRODUCTION

When a new university teacher of CER (Computer Education Research), and this probably also applies to almost all disciplines, is hired they generally proceed with no hesitation in teaching theories to students as the preferred content. They have a PhD, are far ahead of the students in knowledge of the content they will teach, and a theory is the most concentrated form of knowledge. A theory encapsulates an infinite number of cases so that in learning a theory the student is learning an infinite number of cases rather than a limited number of separate facts. This is the best value to the student for their time. Theory is the most value per hour that learning can yield (and similarly the most value for the teacher's time). We all feel that, and we're generally right. So theory seems the highest value from the consumer's viewpoint. So what is theory for? It provides the best content for learners to learn and teachers to teach.

On the other hand if you come across a counterexample to a theory you know about, then that is hugely valuable for a researcher. In a single data point it shows a serious flaw even if it doesn't do anything directly to find a solution. A

Authors' addresses: Steve Draper, steve.draper@glasgow.ac.uk, School of Psychology, University of Glasgow, Glasgow, Scotland, United Kingdom, G12 8QQ; Joseph Maguire, joseph.maguire@glasgow.ac.uk, School of Computing Science, University of Glasgow, Glasgow, Scotland, United Kingdom, G12 8QQ.

counterexample is not a theory at all, but an unplanned observation which is of value not to students but to researchers. Should they be suppressed and unpublished until the observer has solved the problem and so can publish a modified theory to explain them, or would it be better for the field to publish them as they are? These are the kinds of cases, points and latent contradictions which this paper attempts to draw the reader's attention to.

1.1 The aim and objectives of this paper

The overall aim of this paper is to stimulate readers to think more carefully and explicitly about the research activities within CER, and to develop a more thoughtful and active view of the many CER activities and their relationships with each other. It is not our aim to deliver a finished and complete theory – a meta-theory of how CER is and should be done. Not only is that not within our grasp but, as educationalists know, delivering a finished theory is not the best way to stimulate thought since learners then focus on understanding it, not critiquing it and exploring its implications for their own experiences.

This paper's central contribution as a means to that end is to persuade readers that, while theories may be the most valuable outputs of research for those wishing to use them, for researchers themselves there are other kinds of contribution important to progress in the field. This is what relates it to the immediate subject of this special journal issue on theory in CER. We adopt as our criterion for value "contribution to knowledge". By that we mean a contribution to our conscious and explicit knowledge about computing education (even if part of that education process may include some implicit or unconscious concepts and skills of teachers, learners, or researchers). The centre of this paper is a set of 12 categories of research contribution. The set is unlikely to be exhaustive but does constitute a collection of landmarks which together indicate something of the extent of this terrain of differing contributions. As supporting "evidence" each category is illustrated where possible by a CER paper that has been recognised in some way by the community, such as it won an award or is cited widely, in order to keep our judgements in alignment with the actual values of the CER field. We also use examples from other disciplinary areas since the issue of what is a contribution to knowledge is also an issue for other disciplines, and (as expanded on below) much of CER is usefully seen as interdisciplinary. Second in importance as contributions of this paper is that in preparation for presenting our categories, section 2 establishes a set of ground rules for the set of categories. Third in importance is to establish that there is not really such a thing as "the body of CER theory" – not one theory nor one kind of theory is important in CER, but many different types are needed, and increasingly are being used, in CER.

Readers will by now have realised that there are several ways in which this is not a typical CER paper, and that it cannot be written, nor read, (nor valued, nor reasonably rejected) as if it were. You must have a pre-existing successful science before you can have a philosophy of science to explain its success. Similarly any meta-level theory of how CER does or should work is derivative from a pre-existing functioning CER area. These are early days for such an enterprise, and informal arguments seem an appropriate starting point, as well as best suited to engage thought and debate.

Another view is that CER does in fact have considerable debate about this. However it seems that this is almost entirely unpublished and informal: conversations over coffee at conferences, perhaps occasional discussions within research groups. Some conferences (e.g. SIGCSE) have panels that could be about these issues, yet the published record in SIGCSE proceedings for panels has at most 200 word statements of individual positions but no discussion or synthesis at all. The informal tone of this paper aims, in contrast, to be on the printed record but to embody and promote discussion without the air of formal reasoning to a consensus conclusion. It may perhaps be better to read this paper as about questions, not answers – to see each member of our set of 12 categories as a question: "Is this particular category, or is it not, a real and significant contribution to knowledge? And if it is in some disciplines, how is it important in CER?".

1.2 Language, tone, writing style

The style in this paper resembles that of newspaper "op-ed" editorial opinion pieces, which are carefully labelled as such so that readers distinguish them from the factual nature claimed by most of the other articles. Our category of contribution to which this paper itself is closest is "ACE": Argument from Common Experience (§3.3). However perhaps this paper is even more accurately seen as argument from common opinions plus drawing new conclusions from combinations of those beliefs in order to highlight what might be wrong about those opinions as a combination once we notice their implications.

This paper is not presenting new empirical data, nor constructing a new theory by close deduction from accepted premises. This is necessarily so: it addresses the meta-level issue of what kind of work CER does and needs to do; it is not, like most other papers, a direct contribution to CE research itself. This different type of subject matter naturally requires a different treatment. Alternatively it is what Valentine in 2004 [162] (won "best paper" at ICER04) labelled as "philosophy" in his six-category grouping of CER papers, but defined these not as based on authoritative philosophers but as "an attempt to generate a debate on an issue ...". Thus there is after all at least one small precedent in CER for publishing this activity.

1.3 The structure of the presentation in this paper

A different style of argument naturally requires some difference in the structure of the paper. A conventional method section is not appropriate: this is not a paper based on a defined dataset, but more an open-ended set of data points (of cases). A literature review and cited references are differently important here. In a typical paper, one use of citations is to mention past publications as justification for methods used in the study reported: as an authority that will not be questioned nor justified further in the paper. But the appeal in this paper is mostly to parts of the readers' beliefs and experiences, not to some other authority. In particular it questions unpublished assumptions (always more dangerous than explicit assumptions) which of course cannot be cited. Another common use of citations is to mention past similar papers. Within CER, there are few of these, and we have chosen to mention these after presenting our set of categories as a way of discussing our set, rather than as a point of departure. We are presenting this paper as if it were a fresh start, hoping that that will better serve our aim of stimulating thought in readers who should see our arguments as suggestive but not authoritative. It is up to readers to decide what to accept and what to reject because we are not appealing to authority.

The citations in this paper are mainly of the following kinds:

- Citations to published CER or education papers that we use as exemplars of our categories.
- "Further reading" papers – we are grateful to some readers of earlier drafts for contributing to this type in useful ways.
- Citing sources of some of the ideas or points we make, as any scholar should. However these are very often not CER papers nor classified by booksellers as philosophy – but nevertheless have turned out to be sources of ideas to us. They are not used to lend authority, but to show appropriate acknowledgement of the sources of our ideas. Readers will decide for themselves whether, occasionally, to follow them up. In this respect we are entirely in line with standard practice in many disciplines. When you read a CER paper what proportion of the listed references do you then actually read yourself?

1.4 Evidence not from CER

Prior to the definition of categories we define ground rules for the set of categories. These ground rules are not all related, but are all necessary to support the definition of categories. The set of categories does not simply reflect past activity of CER, but is intended to spur thinking about future activity. Consequently, while each category identifies existing CER examples as possible representative candidates of the category, the ground rules do not make significant use of CER literature, but do often use examples outside of CER and across science. The rationale for this approach is to ensure the ground rules act as a strong foundation for forming categories that will better stand the test of time as they are built on common themes observed across multiple disciplines.

1.5 No presupposition about the value of theory

This paper in part critiques any assumption that theory is automatically valuable (see section 5.1), and develops some discussion of the several different uses to which theory in CER is put. However its main aim is to identify the value of many other types of contribution to research in addition to theory.

1.6 The contributions of this paper

The next two sections together present this paper's main contributions, which in order of their importance are:

- A set of 12 categories which together indicate the extent of this terrain (§3).
- A collection of ideas and misconceptions which are drawn on in formulating the ground rules for the set itself as a whole (§2).

2 GROUND RULES FOR THE SET OF CATEGORIES

The aim of this paper is to stimulate discussion amongst CE researchers about the activities in CER. The introduction pointed out that from the viewpoint of teachers and learners, theories are the most efficient form of knowledge, so if they were the only users of knowledge then only theory would be considered and the set would have a single category of theory. If we then consider researchers as a second type of consumer we would have to add some more categories e.g. data collection. However that line of thinking would tend to enumerate only research actions that have been common in CER. That might be done by reviews that look backward to what is already done in CER. However to be more effective at provoking thought we have to think of how to produce a set that isn't limited to what has already been recognised in the past but has some prospect of stimulating new ideas on types of research activity. This section presents the ground rules about how we thought about this in order to create a rounded set of categories which together can indicate some of the breadth and variety of the terrain.

Consequently, the foundation or ground rules for the set of categories are formed from consideration of examples across a variety of subject disciplines rather than only considering CER up to the present. The ground rules are not all related to each other but are all required in mapping the terrain of what contributions could exist and what the CER community might want to get out of any contribution to the field.

The four ground rules for the set of categories are defined and considered over the following subsections.

2.1 Theory is of different types and caters to different audiences

The first ground rule is that there are different types of theory and that they cater to different audiences. Theories may be used by other researchers, but they can also be utilised by other interested consumers. This reality influences the definition of a valuable contribution as authors will want to think about how to disseminate their efforts for a wider audience of consumers. Consequently, we consider the *users of theory and other research contributions* (§2.1.1) and *the variety of types and content of theories* (§2.1.2) as part of the ground rules.

2.1.1 Users of research contributions.

Meleis argues that nursing science requires a close relationship between theory, practice and research that in turn requires a close relationship between theoreticians, practitioners and researchers [95] to address the needs of clients and communities. Meleis argues that despite a shared aim, tensions exist between the different roles within the community in terms of their perceptions of how different users contribute to theory and how they use it. Consequently, in disseminating contributions and/or reviewing them, individuals should consider how they use theory and contributions. This is not dissimilar to Valentine's closing argument that it is important for practitioners to be more scientific in their classroom endeavours [162].

The British Medical Journal (The BMJ), which is a leading medical journal in the United Kingdom with a large impact factor, mostly publishes only articles which report randomised controlled trials (RCTs). Its policies [12] are more carefully stated than that, but are very much aimed at high quality evidence. The reason is that it is consulted by not just researchers but clinicians who want to know what treatments are well supported by evidence. The BMJ has a well justified form of theory-centric policy, as evidenced by: "The BMJ gives priority to articles reporting original, robust research studies that can improve decision making in medical practice, policy, education, or future research and will be important to general medical readers internationally" and "Types of research The BMJ does not usually publish (even if well conducted) owing to lack of usefulness to readers" as well as "We receive many more research articles than we can publish, rejecting around 96% of the research we receive" [12]. The policy is explicitly not applied to other journals in the BMJ family. An interpretation of this policy is that relevant, well evidenced, research findings can be seen as a product to be delivered to those who are not researchers themselves.

2.1.2 The variety of types and content of theory.

As Bertrand Russell pointed out in 1913, Sherin in 2001 [141] did in relation to programming languages and mathematics, and Turing Award winner Judea Pearl [113–116] has recently elucidated in a much deeper way, Physics (and perhaps Science and Social Science in general) has two kinds of theory: ones that specify causal chains, and ones that describe relationships that "explain" or describe (e.g. conservation of momentum) without saying anything about causation. Thus causal relations are *not* a part of some of the most advanced scientific theories. Similarly, much of Sociology is of that kind: describing balances and relationships in society, rather than directly describing what causes what.

Meleis argues a similar position for nursing science, that there are two types of theory: (1) *single domain theories* that are devised to explain and predict phenomena in a given context and (2) *prescriptive theories* that support caregivers in performing appropriate actions and detail potential consequences. Meleis argues these types are developed in different ways, using four potential strategies: (1) Theory to Practice to Theory Strategy, (2) Practice to Theory Strategy, (3) Research to Theory Strategy and (4) Theory to Research to Theory Strategy [95].

The different types and different users of theory directly influence experimental design, methods and motivation. The researcher may focus on causation in experimental design to produce evidence to confirm the importance of a factor in an explanatory theory, whereas a practitioner may focus on causation because they are interested in causing changed outcomes in the world itself.

2.2 The use and value of theory to researchers and other consumers

The second ground rule is about the use and value of theory to researchers and other consumers. This may sound similar to the first ground rule, but is focused more on the consumption of theories whereas the first ground rule was appreciating how they are formed and disseminated.

The production of theory and other contributions to knowledge will take on different forms depending on the nature of a project or experiment. Consequently, valuable contributions need to consider how they are consumed because the needs of a researcher are not those of a designer who wants to put something into use. Consequently, we consider four points in turn: *events have not one but many causes, and not one but many effects* (§2.2.1), *the differences between pure and applied projects* (§2.2.2), *the nature of applications and how they draw on multiple theories not one* (§2.2.3) and *that theory does not always lead to applications* (§2.2.4) as part of the second ground rule.

2.2.1 Events have not one but many causes, and not one but many effects. If the light in the room goes out it might be because someone turned the switch, or the bulb failed, the fuse or circuit breaker tripped, or the local substation shut down. Or, as in 1989, the lights went out for about seven million people in Quebec because there was a coronal mass ejection from the sun that hit the earth's atmosphere and engineers had not grasped that this matters (it was not an issue before long distance cables used such high voltages) [72]. So favourable space weather, or changes in electricity supply to defend against them, turn out to be causal preconditions of domestic lights being on. All events require many causes to be present.

Similarly, events have many effects, even though when humans bring them about they often only care about one of them: the one intended. In almost all ordinary conversation we only think about a single cause and a single effect [30], but in science, engineering, and education we must think more carefully as there will be many hidden assumptions, variables, constraints and requirements.

2.2.2 The differences between pure and applied projects. In conducting pure or applied research, an individual has to decide what causes and effects to focus on and what they want to control. One way of describing the difference between pure and applied research is to say that pure research aims at finding theories that describe a single relationship which is universally true (ideally across all times, places, and contexts), while applied research aims to control all the relationships with significant effects in a single context (a small part of time and space).

Thus, the core logic of pure research is to focus on at most a single cause and all its consequences, while applied research must focus on achieving a single effect and dealing with all its causes (necessary and sufficient conditions). Therefore applied success depends not on one law/factor, but on all the factors with significant effects in the context. On the other hand, in applied research you can ignore true causes when they are small in their effects in the relevant situation.

2.2.3 The nature of applications and how they draw on multiple theories, not one. Following on from the earlier point about every event having multiple effects, we should note that the nature of even simple applications involves multiple requirements which must all be addressed. An important kind of application in education is that of learning designs, or "L-designs"; for example Mazur's Peer Instruction (PI) [93], or Aronson's Jigsaw design[4], or Pair programming[85]. They may have originated in the subject area in question e.g. Pair programming in CER, or in another e.g. PI in Physics and then imported, re-tested and adapted.

All designs have multiple effects not one. All the desired effects are requirements. An important issue, perhaps under-emphasised in CER, is that Intended Learning Outcomes (ILOs) for a class that is using a given L-design are one type (subset) of explicit requirements. The ILOs for the course as a whole, and for making the L-designs employed on the course are two more requirements subsets. Furthermore these latter two sets are frequently not stated yet are important. Discovering tacit ILOs and making them explicit is important in educational research, and is a significant area for useful research. For instance PI often achieves:

- Deeper learning (linking concepts to specific instances and personal experiences) [144, 172];
- Teaching the students to do this kind of linking and thinking as a practice which they come to internalise i.e. they learn how to learn in a way which means they no longer need to have PI organised for them for this purpose after the first semester or two;
- Bonding the class socially by regular peer interaction especially in introductory classes etc. [88].

All of these are desirable, but measuring separately whether each is necessary might also tell us when PI is and is not essential to all those targets, and when it may still be fine but inessential. In fact although "scaffolding" is quite often mentioned in the CER literature, the real educational aim (and closer to its Vygotskian inspiration [163]) should usually be the progressive withdrawal of scaffolding as students become more expert in both discipline knowledge and in knowing how to learn.

2.2.4 Theory does not always lead to applications. There is also the assumption among some that theory is the starting point or that it leads to application. For example, Maxwell's equations for electromagnetism predicted radiation at the then unknown radio wavelengths, which did rapidly lead to the invention of radio communications [135]. On the other hand, arguably the single most important invention of the early industrial revolution was the steam engine. Newcomen's engines were commercially important in mines from 1712 [98]. Watt's major improvement was in 1769 and widely adopted from 1800 [152]. Carnot ("father of thermodynamics") published the theory of how they worked and what limited their efficiency in 1824 [145]. Thus while theory can lead to applications, applications can also lead to fundamental advances in theory. An example for consideration in CER is the application of the Jigsaw learning design. The original design was to reduce hostility, competition and distrust between students [4].

However, the learning design has since been used successfully in other contexts to address other "problems" i.e. requirements; and secondly, to work well in other disciplines e.g. Philosophy, Biology, Psychology [34]. Liao, Griswold and Porter report on the experience of using the Jigsaw learning design in a computer architecture course [83].

The experience report from Liao et al. as well as others in computing education about their experiences of the Jigsaw learning design could, in time, contribute to advancement of theory that has little if anything to do with the original problems of those being addressed by the design.

For the purposes of this paper it is enough to be clear that there are as many cases of applications leading to new theory as there are of theory leading to applications. Even though individual studies generally follow either a pure or an applied logic, large projects often involve multiple studies organised into a research programme (§3.10), and Stokes' main point in his book on "Pasteur's quadrant" [147] is valid for the larger research projects: that both pure and applied goals can be and often are intended and advanced together.

2.3 Science is not mainly about testing theories

The third ground rule is that science is not mainly about testing theories. That is an old view [125], not yet extinct [168]. The misconception is that strong scientific endeavours always start with consideration of theory and while that is true in some cases, it is certainly not true in all cases. Consequently, valuable contributions should not be overlooked because they do not consider theory from the outset. This is not to say theory and other discipline knowledge are not important, they are: but they are not always the starting point for contributions to a field.

Testing theories is one useful activity, but it cannot be the heart of it. Pirsig pointed out that you can generate a hypothesis in a few minutes, but usually it takes weeks at best to test it [122]. Scientific progress could not be possible if testing theories was the whole or even the main story. Somehow there are other processes to generate hypotheses worth testing: to shortlist them.

It is often assumed especially from outside that science is a rational and planned process in itself. While these are valuable virtues when available, nevertheless as will be further discussed again later on, it is important to be aware that unplanned observations are sometimes very important. Almost every space mission to other parts of the solar system turns out to discover something much more important to science than what it had been planned to discover e.g. the Juno mission spotted a volcano on Jupiter's moon Io when everyone had believed that volcanoes could only happen on earth [42]. Thus unexpected and untheorised observations are driving some of the most sophisticated and expensive science today. These are examples of discovering "unknown unknowns": pure exploration. Norris states that out all of the significant scientific discoveries made by the Hubble Space Telescope, only one was originally stated as a goal [103]. Consequently, Norris argues, telescopes that only achieve stated scientific goals are not realising their full scientific productivity. Therefore, discounting unknown unknowns or requiring theory prior to any serious scientific enquiry is to discard contributions to the field in the form of unexpected observations.

For most of CER the perception could be that it should function more like a laboratory science, where theory is the base, variables are controlled and experiments are planned. However, some of CER could function like much of Astronomy, where chance and observation of the unexpected and phenomenon are crucial, if not central. Nevertheless, Pasteur emphasised that "in the fields of observation, chance only favours the mind which is prepared" [111] and as Fabian would argue, scientific progress and discovery in Astronomy is not like buying a lottery ticket [42]. Serendipity is important, but also an individual needs to know what is known and unknown, they need to have a grasp of knowledge and theory, what is relevant and what is important. Consequently, ignorance of theory is not a virtue in CER (nor indeed in many other fields), and a strong grasp of theory is important as it supports unexpected observations and consideration of phenomena. Thus theory is important to progress in multiple ways, but that does not imply that testing theory should be specially prioritised.

2.4 The inherently interdisciplinary nature of CER

The fourth and final ground rule is recognition of the inherently interdisciplinary nature of CER, which is an important point that should support deeper appreciation of valuable contributions and where they come from, but also of what influences their formation. From the beginning CER has been an interdisciplinary field, borrowing methods, concepts and techniques from mathematics, physics, psychology, sociology and many more disciplines as Papert argued [108]. At present, CER remains an interdisciplinary field utilising tools, techniques and insights from other related disciplines. One review of CER [89], especially its section 6.2 and table 1, investigates the degree of borrowing of theories, models and frameworks from other disciplines, calculating that Computing Science, Education, and Psychology are the most frequently drawn upon. It is also important to remember that members of the CER community may be active participants in other domains both within Computing Science, such as Human Computer Interaction (HCI) and Software Engineering (SE), as well as outside it, such as Physics and Psychology.

Fincher and Petre suggest that every interdisciplinary field, including CER, is a trading zone and that the community intellectually borrows from partnering and neighbouring disciplines [45]. They argue that CER participants need to engage with their trading partners, not only from disciplinary neighbours such as the social and learning sciences but also from mathematics and physics as well. The advantage in such trading is that many tools and techniques can be leveraged to investigate a problem, but such diversity is also a challenge as the CER community has to rationalise what is appropriate and acceptable for the field in order to progress [46].

Galison [49] originally proposed the idea of trading zones when rationalising how experimentalists, theorists, and instrumentalists in physics form coalitions to make progress in the field. Galison uses the example of interaction between peasants and landowners in the Cauco Valley in Colombia in terms of money. Landowners have large sugarcane farms with money used to accrue assets and to pay wages. Peasants have a culture of magic and sorcery, and Galison observed that peasants place a different significance on money. Specifically, peasants as godparents hold pesos in their hands when a child is being baptised. The next time the peasant purchases goods, they whisper the child's name with the belief it will return three-fold to the family. Galison argues that while money has a quite different significance in the two communities, this does not prevent both trade and the exchange of money between them, and that brings stability to the local region.

In this sense, trading zones are not only stabilising loops but they can be used as a lens to understand interaction between groups within a community as well as between communities at large. An illuminating example is to contrast the political systems of the United Kingdom and Germany. In the United Kingdom, individual politicians are expected to form a coalition (usually called a "party") and then put forward a unified position to the electorate in advance of elections. Consequently, the primary political parties are each comprised of any number of factions or groups that are likely to differ in many ways on several topics, but are able to find agreement on specific concerns to gain power and make progress. Similarly in Germany, political parties have to form coalitions to gain power and make progress, but this is after the election and the electorate have elected individuals from many different parties.

The concept of trading zones is helpful as it demonstrates that CER should be seen as situated within the context of its neighbouring and partnering fields [47]. More than that the CER community should remember that each CER participant is different and may be trading with any number of disciplines, and may be actively participating in them. Consequently, the tools, techniques and insights they adopt may not only be unfamiliar or anomalous to much of the rest of the CER community, they may sometimes even be offensive as CER extends its set of trading partners.

2.5 Foundation for Categories

Some categories are expected by everyone e.g. theory, and next after that there needs to be a place for evidence collection. Some further categories are needed due to the misconceptions discussed above e.g. the unexpected discovery of counterexamples mentioned in the first part of the introduction, which shows that unplanned but important data can appear and has to be found a place in the set of categories. Finally there are one or two categories which are needed to achieve properties of the set as a whole. You will see that "requirements" is a category whose role in the set is to allow bits of partial knowledge, which connect some categories as intermediate data, to be recognised, stored and potentially published as independent, visible contributions.

3 CATEGORISING CONTRIBUTIONS TO KNOWLEDGE

This section presents the set of categories we have chosen as landmarks to illustrate the extent and diversity of the unknown full set of contributions to research. We present as a table the list of our 12 categories, and then follow that with a description of each category.

There is no single way to order the set. It is likely that this may disturb readers to some degree because they (and we ourselves, and probably almost everyone) are perturbed by several conflicting associations that make us want to think in terms of certain sets or sequences of categories. A later section (§5.2) lists some alternative ways of grouping the categories.

THE CATEGORIES OF CONTRIBUTION		
No.	*Short name*	*Full name*
1	Theory	Theory for its own sake e.g. for explanation
2	Predictions	Theoretical reasoning and predictions
3	ACE	Argument from Common Experience
4	Requirements	Requirements gathering for L-designs at all levels
5	Engineering	Engineering for its own sake. L-designs, patterns.
6	CEPs	Constructionist existence proofs. Research by building things.
7	Planned observing	Experiments, expeditions
8	Unexpected observations	Reporting the unexpected and unexplained
9	Practice papers	Reports on Practitioners' experiences
10	Research Programmes	Sequences of Research studies
11	Reviews	Review papers
12	Tools	Tools to support learners or teachers

Table 1. The categories of contribution.

3.1 Theory for its own sake e.g. for explanation.

This category of contribution "Theory for its own sake" offers explanations and sometimes supports making predictions. In CER a theory may address CS content matter and how to teach it; or address properties of designs (L-designs). Theory may be about *pure* educational theories; or about *applied* matters i.e. generalising about designs (often L-designs) (see §2.2.2). It may also be either home-grown within CER, or imported from other disciplines e.g. Education in general or Sociology, and then applied to CER.

Examples of education theory research include:

- Learning Edge Momentum is a small scale empirical theory, local to CER, to explain data showing bimodal distributions in marks in computer science classes [2]. A paper disconfirming it won a John Henry award at ICER16 [112].
- A paper that draws on brain science and relates spatial skills to student success in STEM in general and CE in particular won a John Henry award at ICER19 [92].
- A paper [84] that applied the SOLO taxonomy (an education theory about learning in any discipline) to CER by analysing the differences between teachers and learners in how they understand program code (from exam papers), won the ITiCSE award of being one the best five papers over the period 1996-2019 [142].
- From Sociology, Tinto took Durkheim's theory of suicide and transposed it into Education as a theory of college dropout / retention [37, 151].

Using the example of The BMJ, we suggested earlier that high quality evidence-sets on applied results were "products" of research of great value to non-researchers (§2.1.1). Similarly theories are often seen as products of research of use to others. We might say that nothing is so useful as a good theory because at their best, theories summarise knowledge that applies not only to particular recorded cases, but to many that have not yet been directly encountered.

The most famous theories, such as Newton's law of gravitation, manage to cover very wide ranges of events. However these are few and far between: many other theories are empirical laws such as Ohm's law, which does not apply to fluorescent tubes or semiconductor devices or superconductors, yet is used everyday by electricians and other work involving regular copper wiring at room temperature. In CS, Moore's law is such an empirical law. Frequently grand theories have only slowly developed from smaller empirical ones. For instance, there were first theories of static electricity, then of magnetism, then of electric currents, then of light; and only finally after about 300 years were all these different patches of knowledge unified in Maxwell's equations. We may expect the same to prove true in Education.

In fact theories may even be vague as well as incomplete, and yet productive. The most basic idea in education theory is perhaps the contrast between Constructivism and Instructivism [31] – between seeing learning as something a learner assembles in their own mind, or as simply an expert telling a learner who passively memorises it. Because there are seriously different ideas on what Constructivism is, it remains a vague concept and yet even as a vague notion it is productive. It remains useful because in courses there is nearly always a lot of simple telling e.g. the names and meanings of basic concepts in a programming language. Yet at any time we can reconsider whether in each course, each class, each exercise getting the learner to construct knowledge would be better overall e.g. by getting learners to identify and learn the missing concept in what they have been told so far, or by asking them that if a new concept were to be added to a programming language what should it be and why, or by getting learners to teach their classmates about each concept (because trying to teach someone else is a powerful boost to your own learning).

Within CER, [90] reviews the theories currently being used there. Education theories may be and often are tested and researched across all disciplines, while some are more restricted in scope. They may investigate whether the theory applies in all countries, or only some. Even what it means to learn may vary depending on culture: western cultures almost invariably use ability to recall as the way to judge learning, whereas some other cultures judge learning by ability to recognise [23]. Because education is organised nationally, its nature may be very different in different places. In countries where education is not compulsory, for example, schools seldom have to worry about whether learners are motivated to learn nor whether they are capable enough to benefit. So in fact no education theory (or CER theory) can be reasonably accepted until it has been separately tested in many countries.

The above has looked at *pure* research (§2.2.2) on education theories, whose common feature is that their truth and scope is researched and studied, not how to do teaching and learning in specific ways, although of course education theories may inspire applied work on creating novel L-designs. In contrast, we now turn to *applied* theories (§2.2.3) that aim to generalise about applications which are successful pieces of learning and teaching (e.g. L-designs), and each of which involves many requirements, perhaps related to various education theories. Examples of research on theories about education applications are:

- Sanders et al. review the use of the term "Active Learning" in the CS related literature [133]. They found however that very few defined the term, that it was almost always used to refer to teaching (not learning) activities, that it always meant "good teaching", and they selected four dimensions in which to arrange the various methods:
  - How active were the learners and in what way?
  - How reflective?
  - How social?
  - How specific were the methods to computing science?
- In contrast Hake [53] in the field of physics teaching had in 1998 short-circuited this by asking each teacher in his survey whether in their opinion their teaching approach made substantial use of "interactive engagement"

to promote understanding of the content to be learned, and saying that interactive engagement meant learner activity (perhaps with material) with immediate feedback from teachers or peers. He reported a statistically significant difference with a large effect size between those who did and did not. This shows that there can be effective ways around the issues of definition, since the 62 instructors who responded were not pre-trained in the definitions or theory of "interactive engagement" or "active learning".

- Chi's ICAP framework [20] was based on numerous experiments (both in the lab and in the classroom), where she showed that a number of different types of "Active Learning" (different members of one family of L-designs) may be ordered by their effectiveness i.e. some forms are reliably more beneficial than others. It is essentially a theory of peer interaction and its power, under the right conditions, to rival expert skilled teaching by an adult tutoring in groups of 2 or 3 learners. Her experiments were on school maths tutorials, but seem likely to apply to CER; and also to suggest that the method could be applied across classes of all sizes because she showed that video recordings of a tutorial were about as good as with a live, face to face tutor.
- The L-design Peer Instruction (PI) has been the subject of post-invention theorising about how and why it works [35].

Much (but not all) applied education research however concerns L-designs, their invention, testing and generalisation. Good L-designs often seem to have a life of their own and have properties that their inventors did not expect or understand fully. This has been recognised elsewhere as due to "the task-artifact cycle" [19] in which the originator has an idea for something useful, creates and distributes it, then finds that users quickly find new uses for it. The first spreadsheet was designed for doing accounts (recording past transactions), yet the first population of users contained many who used it for planning (reasoning back from fixed totals of budgets to look at what they might spend it on). In recent times, applications like Facebook take this much further, constantly studying how their invention is used and modifying it try to serve emerging needs better: often needs that no-one had before. (This of course completely contradicted old top-down approaches of first eliciting requirements, then implementing them and only them.) Furthermore, successful L-designs often satisfy implicit requirements such as being activities that are easy to explain to students, that make a class bond together while learning together e.g. [88].

Consequently, a classic example of applied research is to discover whether an L-design originally invented for teaching one subject also succeeds in another discipline (as many do) for instance PI in physics [25], and in Genetics [144]. This is useful in practice, and may also lead to understanding why some transfer like this, which would amount to a more general Engineering theory within the area of educational applications.

To get some idea of the very large probable scope for future theory in CER, consider "micro-nutrients" in nutrition. Besides our bulk needs to consume water, calories, protein and some fats, there are about 28 vitamins and minerals essential to health. Remembering that all events need multiple causes to be in place, we may expect that there are many more essential conditions for learning in education than in nutrition, since life has always depended on nutrition whereas it does not depend on formal education. Each such cause is likely to need its own theory.

3.2 Theoretical reasoning and predictions.

To take a theory and then work out what it would predict in some new particular situation, and hence prepare the way for new empirical work on that situation, is a distinct research step. As such it is a contribution to knowledge, and should be recognised as such. On occasion, perhaps it should publishable as a paper in itself.

In disciplines very different from CER this has sometimes been famously important. Eugene Parker submitted a paper to “The astrophysics journal” showing mathematically how there could be a large solar wind of particles, but it was rejected by the two referees on the grounds that it was absurd. Asked by the journal editor whether he wanted to publish it, he replied that the referees had not offered any substantive criticism so yes he did, and the editor published it in 1958 [109]. In 1962 the Mariner 2 spacecraft, launched primarily to study Venus, returned data showing there is indeed a strong solar wind of charged particles flowing out from the sun, and that Parker had been right. Parker later became the first person to have a spacecraft named after him while still alive largely because of all the subsequent work that flowed from the phenomenon, for which he had provided the theoretical analysis in advance. We know the details mentioned above because years later Parker wrote an amusing paper that is well worth reading “The martial art of scientific publication” [110] about the peer review process, including these details (and also cases when the referees had been right and himself wrong).

Is this type of contribution important in CER? One example is a paper by Kallia and Cutts [70] which won “best paper” at ICER21. However this paper does both substantial theoretical reasoning work (§3.2) by applying Bourdieu’s theory from Sociology to CER, and additionally a systematic literature review (§3.11) and also a thematic analysis (§3.7). In other words, it uses the considerable amount of empirical work already published to point out and explain the limitations of its application of theoretical reasoning, thus combining three types of contribution.

Another example where the theoretical reasoning was done but *not* published separately is in the work by Tshukudu published in a series of seven papers including [153, 156] the last of which [156] was “noted as a runner-up to the Best Paper” at Koli Calling 2021. (These are also discussed below under the “Programme of research” category, see §3.10.) This series of papers presents a sequence of studies looking at the kinds of errors that learners make when learning either their first or second programming language. The first step in the argument in each of these papers including the first was to present a way of thinking about transfers from learners’ prior knowledge of either their natural language (e.g. English) or of a previous programming language. To do this Tshukudu drew on one theory of the problems programmers show when transferring to their first object oriented language, and another theory of the problems people show when learning a second natural language. The theoretical reasoning of how these two theories might apply to the new domain of students learning programming is original, but was not published and discussed by itself, and is re-presented in nearly every paper.

What other work of this type could be imagined in CER? Two examples could be to identify a theory in Sociology (e.g. about effects of race or class) and ask what they mean in educational situations, or theories from Nutritional Science and ask what they mean in education (hungry children do not learn effectively).

A third example that would be solely a theoretical reasoning contribution and might be productive in CER would be to present contradictory predictions about a situation, thus clearing the ground for and motivating a study that tests which theory applies and when. For instance Higher Education teachers frequently complain that students won’t do any suggested work if it doesn’t carry marks (contributions to grades). Yet when supervised in a class (e.g. as in Mazur’s PI) students seldom resist voting on a question, or discussing which the right answer is with other students. (And in this case, giving credit for voting can reduce the benefit according to [126].) In yet other cases students no longer need to do this thinking in class or even with peers, but develop a habit of thought of thinking about the pros and cons of a proposed concept, or thinking of concrete examples of it. Perhaps students come in time to carry out activities because they have learned they are useful regardless of what the teacher says: a third theory of "motivation" less often mentioned by teachers, but which if true, carries implications for whether and how we ensure that our students do learn this reason for important activities.

More examples can be found in a 2019 review [90] in its section 4.4.2 which lists half a dozen uses of theory in CER to make predictions. Such work addresses Pirsig's [122] point that random hypotheses are cheap to generate compared to the empirical work needed to test them and so cannot be the driving force behind productive science. Discussing theoretical reasoning could allow theories and their applications to be partially debugged before investing in empirical work. Taking a general theory, and making specific predictions about a new kind of situation takes intellectual work, and is a significant contribution in itself, prior to and separate from doing any empirical work. It might be that publishing and discussing such work in its own right would accelerate the field of CER.

3.3 ACE: Arguments from Common Experience.

While research papers normally deal with theories and new data, there are papers that do neither but simply make arguments based on things all the readers accept as part of their shared, common experience. The opening two sections of this paper are of largely this kind. The whole of this paper mostly belongs in the ACE category, but because we also use references to published papers to illustrate and support many parts of our argument, it also has some aspects of a Review (another category in this set, see §3.11).

Sfard's paper on two metaphors for learning is an argument based on analysing the language we use about learning (English in the case of her paper), and this is common to all readers [137]. Yet from it she concludes that it shows a fundamental limit to the coherence of our educational theories. Her two metaphors, the acquisition metaphor and the participation metaphor, show that part of our talk is about having definite states of knowledge which we can attain and pass on, but other parts are about participation and doing not knowing.

Arguing from common experience is pervasive in many parts of philosophy, going back to Socrates. It is, or originally was, the justification for Euclid's postulates and deductions: the rest of Euclidean geometry could be proved if the reader agreed that these postulates were self-evident. If the arguments can be supported by things all readers accept then there is no need for new data or other evidence. In some disciplines there are venues for stimulating a debate led by a paper which presents issues and arguments, with new and wider suggested implications but not new data: for instance the journal "Brain and Behavioral Sciences" (see [79] for an example) which have some of the flavour of ACE.

Although still a rare mode of contribution in science, there is another, but strong, form of this type of contribution in which experiences are created which almost everyone shares. Such demonstrations do not need experiments nor statistics on how many participants were used, and yet are much more convincing. Adelbert Ames was a painter among other things who turned to research in visual perception and advocated this approach. He designed and built about 50 such demonstrations, and many readers, though they may not remember the name "Ames", will have seen an Ames Room in science museums or exploratories where you look through a peephole into a room where people's size seems to vary depending on where they move in the room. (This is a demonstration – and so creates a shared experience – of how, when stereopsis and binocular vision are disabled, other cues of depth perception such as the expectation of a rectangular room control our perceptions [62, 63].)

Similarly lecture demonstrations at their best may be of this type. If in a lecture on peristalsis, the lecturer performs a handstand on the bench and remains upside down while drinking a glass of water through a straw, you cannot go on believing that food gets from our mouths to our stomachs only because we eat when upright and gravity makes it fall downwards. Therefore, this mode of arguing from common experience has direct educational applications.

Arguments and actions based on common experience is of potential importance in CER. Part of learning to program is seeing what happens when your code is run or interpreted: a highly personal experience (although the strength of the impact of such repeated experiments whose input you directly controlled depends on how much you put into the exact code you ran and how strongly you predicted its effects). This shared personal common experience is not present in most disciplines unless they not only have lab classes but ones where the student controls the conditions tried. It is common for CS classes and courses to insist on all learners learning to actually create functioning code. It is common for such courses to believe that this enjoyment is important e.g. with engaging graphics, simulations, using block-based languages. These are the sort of conclusions many individual teachers have reached but have less often articulated or tested. It suits many students to identify success in CS with the gratification of creating something that "works" rather than with understanding the exact meaning of code. Dziallas and Fincher illustrate that some programmers may be many years into a career as a programmer before realising that immediate utility and relevance are not the only way of judging what it is useful to know [38]. This is a case of learners using ACE to justify their view of what matters in programming over any other view that the teachers may have. However conversely, similar reasoning about the common experience of enjoying the personal power that programming affords is made by teachers and some researchers, such that they choose to teach languages that enhance that enjoyment at the expense of focussing on the exact meaning of code. Thus arguments from common experience are frequent and influential in the practice of CE but if they were printed and critiqued this might test their validity more often.

3.4 Requirements gathering for L-Designs at all levels/scales.

Engineers of all kinds are familiar with how requirements gathering is the first stage of design. It may be of especial importance in education because successful L-designs tend to have many implicit properties and identifying these is one way of making knowledge about educational activities and designs explicit, and then communicated as a research activity in itself, thus advancing practical knowledge in CER and education (see §2.2.3). Here we will alternate between two kinds of example: the civil engineering of bridges (because of our debt to Petroski's ideas [120, 121]) and CER.

L-designs are an applied enterprise defined by being something created by people to have a specific effect on learning (the original requirements). On the one hand many causes (which are all requirements) have to be in place to create the desired effect, and on the other hand any solution (design) will have multiple effects some of which may be undesirable, and avoiding each of these must be added to the list of requirements (§2.2.1). Thus the "client's" original requirements may soon be expanded by an expert designer who foresees their consequences and adds these as implied negative requirements.

In civil engineering it is not enough to satisfy the requirements of a bridge as it will be after it is built (e.g. the load it can carry, the winds it can resist), but also how it will be built and how it will be maintained. In PI (peer instruction) you have to have all members of each student group communicating with each other at the same time: entailing requirements of being active at the same time, and having access to equipment or space to do this; and normally also having the whole class interacting at the same time and with voting equipment. These extra entailed requirements mean that PI cannot easily be part of a self-study online course – summarised as an implied negative requirement of not being part of a self-study course.

Petroski has argued that there is a further potential problem due to automatically satisfied requirements which may remain unrecognised indefinitely [120]. This is because there are unlimited numbers of implicit requirements, but it is not necessary to know them as long as they are automatically satisfied by the available methods. For example, stone bridges such as the Romans built will never be blown over by any wind on the surface of the earth because by the time they are sturdy enough to support the vertical weight of the bridge (and traffic) they are too heavy for the wind to affect them. But as materials allow bigger, longer, and lighter bridges at some point a new requirement of addressing side winds makes itself apparent. The example Petroski uses is the collapse of the Tacoma Narrows bridge. While there have been disputes about what exactly the requirement and / or the solution is, the main point remains: that some requirements are discovered only through disasters, although finding what theories best address this may take time. The solution that emerged after a time was to treat the decks of bridges as aerodynamically like aircraft wings, but to design them like inverted wings, so that higher winds increase the down force thus stiffening their resistance [121, 136]. The collapse was an example of *unexpected observations* (§3.8), and it was a counterexample to what engineers had thought they knew. Part of the solution process was adopting the regular use of wind tunnels (only just becoming available at that time) to test scale models, thus allowing *planned observations* (§3.7), and which are also useful when your best theories are not fully adequate.

Generally education is quite backward in developing explicit requirements, and there may be considerable potential for future improvements here. In education Intended Learning Outcomes (ILOs) are often thought of as the only requirements specified and focused on, and are very often only about course content ("c-ILOs"). However there is no real boundary between content announced and assessed, and unassessed but clearly desired aims which are also learned e.g. graduate attributes such as adaptability, resourcefulness, and communication skill; and learning how (better) to learn in the discipline (e.g. in CS, learning how to teach yourself a new programming language is widely valued by employers and generally realised by senior staff but seldom directly stated). In addition, a department might have ILOs for the year as a whole and for the degree programme as a whole, but not communicate these in a way that an individual teacher thinks of when they design a module. Dziallas and Fincher opened up an unusual angle on c-ILOs by illustrating how some of the value of their original degree in CS was understood by graduates only much later after some years in the industry [38]. This could be used to inform c-ILOs at the degree programme level. Full c-ILOs for a module should really say how that module contributes to these larger scale ILOs. Furthermore, there are other kinds of intended outcomes such as eliciting interest, and "Social ILOs" such as creating group cohesion [88]. Essentially some commonly adopted L-designs have multiple effects, and in fact multiple desirable effects, not all of which are originally planned for, nor well evidenced when they are noticed. This might be called inverse requirements gathering: using evidence of positive side-effects to identify good effects as new requirements to adopt.

For example, PI clearly earns substantial approval from learners (e.g. in the form of recommending its adoption in more courses [80]); and it produces deep as opposed to shallow learning [144]. Doing inverse requirements gathering by identifying these effects is a way of addressing the task-artifact cycle [19] and publishing them as, say, "satisfied requirements" with respect to a particular L-design will help practitioners in selecting an L-design, and linking it to their courses' requirements. Equally it will suggest that studies of a whole course or of one L-design within it should find a way to measure how well each requirement was satisfied – and so support the design of future research.

Beryl Plimmer produced a tool to help her mark her first year programming assignments [123] (also discussed in §3.12). She used task analysis to produce a set of requirements for the tool, and after several iterations on its design, ran an experiment that showed marking with her tool was both faster for the teacher and better quality from a learner viewpoint than doing it on paper. It is a case of making the experience of an experienced teacher explicit, visible, and understood for the first time. In turn the power of doing this became apparent by showing that the superficially very different case of marking essays could be supported in the same way. In both cases there are multiple bits of paper that need to be in view at the same time. On the "input" side for programming: code, test outputs, ...; for essays: the exam question, the meaning of the grades on the university scale. On the "output" side, at least for some essay marking, the places where the marker records information:

- The place where the mark is recorded and sent to the teaching admin,

- The place(s) where feedback to the student is recorded (e.g. on the script; and on a form or other message),
- Notes to self on reasons for marks for use if marks have to be reconciled between markers,
- Occasional notes to self about ideas and references the student expressed which the marker wants to look up later.

This is an example of how serious requirements analysis (task analysis) is illuminating of both an actual design but also of the problem in general, and of how an interview alone would be unlikely to have elicited these requirements and that observing the work is important. It is also an example of how supporting teachers can be as important as supporting learners. And it is furthermore an example of how good analysis can sometimes surprisingly generalise between different disciplines and task details, even though the actual tool would not directly transfer since much of its value is in the details of what input and output documents have to be assembled in each case of assessment, which is why this belongs under "Requirements" and not "Tools".

3.5 Engineering work for its own sake. Learning designs and patterns.

Engineering work for its own sake, both design and construction, has its own direct importance across all engineering areas, as the examples discussed for *constructionist existence proofs* show (§3.6). In education generally, including CER, the most important products of applied research (§2.2.2) are probably L-designs. This is different from the category of *Theory for its own sake* (§3.1) because the output here is an instantiated product that works materially, rather than an intellectual idea that works for reasoning in some area.

Authors of L-designs usually do not have a good and complete theory of why and how the design works – that would be theoretical work (*Theory for its own sake* §3.1) on why a design works, not design and construction work (this category). It is important and interesting that some L-designs turn out to work well in quite other disciplines and situations than the ones they were originally devised for, and sometimes to satisfy different requirements from the ones they were invented to satisfy.

Some examples are:

- There is an L-design in which students are required to create Multiple Choice Questions (MCQs) for their classmates. This is one type of "contributing student pedagogy" [24, 54, 55]. This is effective educationally because writing a *good* question requires you to understand a topic much better than other topics. Denny et al. illustrates how their software, PeerWise, supported this activity and applied it in real courses and reported on that (an experience report) [27]. Denny et al. also collected and reported quantitative data that supported the claim that learners using PeerWise engaged in deep learning, and performed better in exams, and so that there was a real educational gain because of it. The paper won a John Henry award at ICER08.
- Contributing student pedagogy is a *whole class of L-designs*, and a paper by a working group reviewing this class as a whole [54] was declared as the Top ITiCSE WG Report for the period 1996-2019 [142].
- Peer Instruction (PI) [25, 93] has been widely used in a number of STEM subjects including Physics, Genetics [144], and Computing Science [172].
- Jigsaw is an important and strong form of student-generated teaching, invented by Aronson et al. in 1971 [4], and described in more detail in section 2.2.4 [34]. It has since been used successfully in quite other contexts to address other "problems" i.e. requirements; and secondly, to work well in other disciplines e.g. Philosophy, Biology, Psychology. It has also been used in teaching some CS subjects e.g. ethics [86].

In education generally including CER the most important types of applications research are probably:

- The invention of new L-designs for old situations (traditional subjects and teaching methods), and collecting evidence on how effective each is.
- Designing and implementing new L-designs for new situations.
- Investigating whether an L-design originally invented for teaching one subject also succeeds in other contexts and disciplines (as many do). Trying and documenting these re-applications to new domains is useful research, showing that transfer both into and out of CER may be as valuable as raising confidence in the L-design's success over differing contexts.
- Showing that an L-design satisfies different requirements from the ones it was invented to satisfy.
- Identifying new, unarticulated properties of old L-designs e.g. tacit requirements now identified, and measured as satisfied in old L-designs.

L-designs are where research is most likely to observe the importance in education of not just explicit conceptual knowledge but also of implicit skills. Crouch and Mazur's paper [25] reports on the learning gains due to the introduction of Mazur's Peer Instruction (PI) L-design in his own class. It shows the performance the year before introducing PI, and then the year by year performance of PI over the next seven years. The bar chart shows an immediate doubling of the amount learned in the first year it was introduced, but then steady additional annual improvements until in the seventh year of PI about the same amount of increase again had been attained. This is a direct demonstration that the concept or explicit idea of the L-design achieved a big gain in one lump, but that the same amount again is gained by improvements in the instructors' delivery of the L-design (perhaps partly by increasing tacit skills). This shows that teacher skills are an important factor in learners' degree of success: a conclusion also drawn from quite different data by Chingos [21] on school teachers. These skills are probably implicit or at any rate tacit (not spoken about) because they appear to come from experience and not from being taught by one teacher to another.

PI itself also works due to the learners' tacit skills or practices. The students are made aware of their own initial opinion by the first vote; then aware that others' have different opinions by either seeing the distribution of votes or by the peer discussion. They find discussion about reasons natural to do in those circumstances, without knowing or believing that this is an important learning activity. After two semesters of this practice they may develop different (perhaps also tacit) habits if they have by then internalised the practice of asking themselves why they should accept an idea given them, or to run it over in their heads to see how it connects or seems to be inconsistent with other facts and experiences they have.

Implicit learner skills are also important in programming. A lot of maths for STEM subjects does not just teach the idea of each concept or operation but gives copious practice through simple exercises, because to be useful the learner must be well practised enough that using the technique does not require slow recall. In learning a programming language the same applies: it is not sufficient to be able to recite a definition of a syntax element such as an if statement. In reality the student must learn the common "patterns" in which it is used so that they can compose new code fast enough to be useful even at a student level. Knowledge given infinite time to recall and reconstruct it is not enough: fluency and skill are in fact required, whether or not either teacher or learner recognises and states this.

3.6 Constructionist Existence Proofs (CEPs). Research by building things.

Existence proofs are those where discovering a single counterexample defeats a proposition. An old example (invented by philosophers) is to consider a male European naturalist who has only seen or heard of white swans, but has to change his ideas and definitions of what a swan is as soon as he goes to Australia and sees a black swan. When the first, rather poorly preserved, specimens of dead platypuses were sent to Europe, they were thought to be frauds because they seemed to be cobbled together parts of several unrelated animals. However when live specimens were sent back or seen by trained naturalists, the accusations of fraud faltered and theories had to be revised: the existence proof was decisive.

In Education, L-Designs are arguably the most influential publications e.g. PI (Mazur's peer instruction) [25, 93], Reciprocal Peer Critiquing, the Jigsaw classroom [4], Patchwork Text, etc. In CER, PI [93] and Peer Assisted Learning have been used, and also Pair Programming [85] and still other L-Designs. None of these was fully understood before being introduced, but they drive practice forward along with increasing understanding of how and why they work, and of the conditions in which they do and don't work well.

In Engineering, constructions carry a conviction all their own, regardless of any theory or explanations the engineer may give. All applied disciplines including Engineering, Computing Science, Education, and Medicine have a prominent place for what we will call "constructionist existence proofs", where the essential contribution is a design that was invented, built and tested. While it is usual to accompany these with descriptions of a rationale behind the design, their biggest contribution to knowledge is that the design is now known to work in at least one case, and so can be replicated even if that would entail rediscovering the techniques.

When Russia orbited the Sputnik, non-experts could immediately grasp that they might now drop things where they liked on earth, and also point cameras where they liked. When you saw or read of the Spears tower in Chicago, you knew people could now build things over a quarter mile tall. When you hear of the Burj Khalifa, which is just over a half mile tall, you are even more impressed, while knowing almost nothing of the problems the engineers can now conquer (e.g. that it is built on shifting sand dunes, and underneath them is an especially weak and crumbly rock; and that it needs sophisticated theories and solutions to the effects of wind on the building).

When the Wright Brothers demonstrated controlled, powered flight for the first time in 1903 the goal of flight went from a lengthy history of amateur failed attempts to a reality ripe for intensive development. Only 11 years later (August 1914) the Western allies had 30 aircraft in active military service in the opening campaign of Word War I. It was the proof that it could work that drove this great change in pace, and this rapid progress was driven not by publishing or testing theories, but by building a series of working demonstrations. Constructionist existence proofs (i.e. novel working designs) in fact generate research questions more than answer them.

A more sophisticated use of CEPs was demonstrated in a paper by Bloom [11]. He measured the learning attainment of average children with a skilled tutor in 1:1 teaching. This is economically unaffordable in mass schooling, but it demonstrates what is humanly possible to achieve; and thus demonstrates by existence what can be achieved. In his paper Bloom then discussed how he planned to direct his research to find L-designs which when combined can reach this human limit, but affordably.

In one way CEP is related to ACE as a contribution to knowledge, because it is a form of argument that is largely to do with reports of something which is taken as common experience, and which many have seen with their own eyes. In another way, it is related to Engineering work for its own sake: getting something to work, which then has its own power to drive research (both practical and theoretical) forward independently of prior theoretical underpinning.

3.7 Planned Observing by and for researchers. Experiments, expeditions.

Planned observations cover a wide field, and include qualitative as well as quantitative data. For example this ICER16 paper [38], in an exemplary methodology section, outlines the different types of qualitative study, why they selected the one they did, and what it consists of.

Planned observations aimed at testing or disproving theories belong in this category. Zingaro and Porter [172] tests the theory that students learn from Mazur's Peer Instruction (PI) method in the domain of computing science, and that this is not due to memorising answers but to understanding at a deeper (transfer) level . It builds on [144] which demonstrated this in a first year Genetics classes, and did so by introducing the use of "isomorphic questions" (published in the journal *Science*: current impact factor 42). A different kind of theory-testing educational research is [74] on revision methods for Higher Education (HE) students i.e. this applies in CER too. Its theory type is the cognitive psychology of human memory. The study showed that repeated testing of recall is not only much more effective than re-reading notes for rote recall, but is also better when the test is on the meaning of the material; and it is better than revision requiring elaboration of the material instead of repeated testing. Methodology papers (describing or comparing methods of observing including experiments) belong to this category as well.

Finally "expeditions" are fully planned as to the actions and often as to the measures to be recorded, but with no definite plans as to what will be discovered. The historical Marco Polo's expedition was of this kind: setting out to discover the practicality of a potential trade route (was it survivable) and what goods might eventually be traded. These have become ever more important in science. From sending plant and animal collectors, where the place was planned but not the undiscovered animals (obviously), to discovering undersea vents ("smokers"), planetary space missions where the planet was planned but as it turned out not the unexpected discoveries, to huge space telescopes mapping all the stars in one direction where the instrument is planned but the discoveries are about patterns of movement in huge sets of stars.

Thus planned observing includes both testing theory where the observations are usually as expected, and data gathering where the observing was planned but the observations were not, and unspecified theories are used only post hoc to notice surprises that were not predicted. In education the work of Sugata Mitra may be an exemplar where he has demonstrated in various ways how much children can learn by themselves without teachers, which he has reported mainly in videoed talks, rather than in journal articles.

3.8 Unexpected Observations. Reporting the unexpected and unexplained.

Unexpected Observation papers report the unexpected (rather than exploring the expectations of theorists). These are of great importance in science. In Biology there is no predictive theory of what species exist, and discoveries are still being made. Humans are classified as great apes, and in 2017 a new great ape was discovered bringing the total to eight. Estimates are that there are 300 undiscovered mammals, and that only one in every 100,000 species of bacterium have been discovered, including about 2,000 new ones very recently discovered in the human gut. Space missions exploring the solar system are selected on the basis of expected new findings; yet almost every mission turns out to discover something much more important to science than what it had been planned to discover e.g. the Juno mission spotted a volcano on Jupiter's moon Io when everyone believed that volcanos could only happen on earth. If you look at textbooks on testing (in Computer Science, in engineering, ...) they usually only discuss tests that will detect frequently found problems. Much more important to research (as opposed to quality control in manufacturing) are the tests that discover "unknown unknowns": this is what unexpected observation papers report.

An example from Physics Education Research is [118] which reported that when each student's seat in a lecture theatre was randomly assigned, then their seating position in the first half of the semester statistically significantly affected their final grade in the course. (The course was for physics, but there is reason to expect this to be a general result across disciplines including CER.) The researchers had expected there to be an effect that reversed when student positions were reversed half way through. They therefore reported a dis-confirmation of their prediction, plus the factual observation of a different effect which shows that a different theory must be found.

Observations are on a spectrum from wholly unplanned through to planned; with an important middling set ("expeditions") where the measures and observing were planned, but the important observations / results were not expected yet noticed. The expertise of the observer makes a big difference to noticing the unexpected. Alexander Fleming noticed something unusual when he discovered penicillin: but few others would have noticed that the pattern on a bunch of agar plates left uncleaned before a long weekend was unusual. A personal example of this happened when colleagues were chatting about feedback to students, and eventually turned to a student who was patiently waiting and asked her whether she thought the feedback the department provided was useful. She replied that all she wanted to know was whether she could and should continue with the subject next year. This showed that the learner's view of feedback was quite different from the staff's teacher-centric view, because learner goals are different, and led eventually to writing a paper exploring this [33]. This is a classic example where no observing had been planned but theory-related expectations caused "noticing" like Fleming's.

3.9 Reports on Practitioners' experiences.

Practice papers may contain few observations, yet depending on the eye of the reader, may still alert someone who sees significance in an unplanned observation. But equally many papers may qualify as practice papers but also in other categories e.g. the Cutts et al. concept of a Thinkathon [26] which was an ad hoc practice intervention although by adding extra observations it turned a one-off event into something more; and Guzdial's paper [51] on a progression from pure practice to something more. (Both these papers, discussed also in other sections, won CER awards.)

In one way practice papers are a kind of expedition paper: what the teacher did and how they felt about it. But the historical Marco Polo brought back trade goods, and an account of his survival and route. Perhaps practitioner reports might be usefully enhanced by how the teacher felt compared to other courses they have taught; or include a report by an observer of what the teacher did (which would certainly raise the chances of implicit skill and knowledge being reported). In fact a big barrier to L-designs being usefully adopted is that teachers tend NOT to follow an L-design but to adopt bits that are convenient without regard to what is important. That cuts both ways: descriptions of L-designs frequently do not specify exactly which bits are crucial i.e. their requirements for execution have not been carefully reported. In fact there is a literature on how very important teacher quality is – and that is probably almost all about the tacit skills a teacher does or doesn't have. Contrary to many assumptions, as measured by student learning outcomes in schools, it is more important whether they get the best or worst teacher in a school, than whether they go to the best or worst school in the town [21].

3.10 Research programmes: Sequences of research studies.

Research is frequently organised into a sequence of stages of different types of contribution, not a single grand experiment. If you have a theory you might construct a hypothesis and do an experiment. If you are doing an accident report, you have the effects and try to work back to the causes. In Medical trials there is a more elaborate sequence: lab trials on animals, then three "phases" of testing on humans. The UK Medical Research Council offers this model on p.4 of [99]. The model offers essentially this:

(1) Pre-clinical. Explore relevant theory to ensure best choice of intervention and hypothesis and to predict major confounds and strategic design issues.

(2) Modelling - Phase 1. Identify the components of the intervention, and the underlying mechanisms by which they will influence outcomes to provide evidence that you can predict how they relate to and interact with each other.
(3) Exploratory trial - Phase 2. Describe the constant and variable components of a replicable intervention AND a feasible protocol for comparing the intervention to an appropriate alternative.
(4) Definitive RCT (randomised control trial) - Phase 3. Compare a fully-defined intervention to an appropriate alternative using a protocol that is theoretically defensible, reproducible and adequately controlled, in a study with appropriate statistical power.
(5) Long-term implementation - Phase 4. Determine whether others can reliably replicate your intervention and results in uncontrolled settings over the long term.

Shayer gives this model for applied educational research on pp.112-3 of [139]:

(1) Primary effect study: what effects, and how large, can be achieved using the intervention? Is there an effect (whatever the cause), what effect, what size of effect. In this stage you do an experiment and show there really is an effect, defensible against all worries. But you still haven't shown what it is caused by.
(2) Replication study: can it be done by other enthusiasts (not only by the original researcher)? Can the effect be transferred from the researchers to any other teacher(s)? (Or is it only achievable by the originator and therefore may depend or even be wholly due to some unconscious skill on their part?) The intervention at this point is probably still using a highly self-selected and unusually enthusiastic and able teacher; but at least it has been transferred. Here you show it can be done by one or more others: so at least it is not just a charisma effect, or the implicit skill of the researcher as a teacher.
(3) Generalisability study: Can it be done by non-enthusiasts? i.e. can it be transferred via training to the general population of teachers without special enthusiasm or skills? This is actually a test of the training procedure, not of the effect — but that is a vital part of whether the effect can be of practical use. Can a teacher training course be created for transferring the intervention, and is it successful in achieving the effects for pupil attainment even with perhaps reluctant teachers? This is obviously essential for the research to have any significant effect nationally. Here you are testing whether training can be done.

Ann Brown also has interesting comments on stages of educational projects [17]. Less generally, one might see Adey & Shayer's own CASE work in middle school education in multiple disciplines [1] as having an extended number of stages.

(1) Theoretical work.
(2) Development of materials derived from the theory, to use in the next stage.
(3) Primary effect study. Do these materials benefit learners in a special study outside the school timetable and constraints?
(4) Replication study: can it transfer to any other teacher(s)?
(5) Generalisability: create, develop, and test a teacher training course.
(6) Roll-out in multiple schools (test that it generalises over institutions as well as teachers).
(7) Follow-up studies on long term effects.

Although relatively uncommon, one type of educational research, therefore, is a connected programme going, for example, from theory to successful applications of it. It is important to recognise this type for two reasons: they let us glimpse explicit reasoning about how to choose what research to do, and different steps in a programme generally require markedly different types of research (as opposed to a common productive pattern with some researchers of using just one method used repeatedly in different situations).

The work by Tshukudu on the topic of problems learners have in learning their first and second programming languages (which was also discussed for a different reason above under the category *theoretical reasoning and predictions* §3.2) is an example of a sequence of seven papers with different kinds of study building on each other: [5, 153–158] which together show something like the following stages:

- Identifying an unexpected problem needing research [5].
- Develop the application of two theories from natural language learning to learning new programming languages [153].
- Exploratory qualitative study of the phenomena in her theory, using repeated interviews of five students transitioning from a first course in procedural Python programming to a second course in Java, and the carryover of concepts between the languages [154].
- Tested the theory by looking at problems of two predicted kinds and showing that the data verified that both their occurrence and their relative difficulty for learners matched the theory [155].
- Piloted one element in the future L-design ("explicit instruction"). A study on how interventions by human tutors affected difficulties. Showed that it helped in one type of difficulty more than the other as predicted [158].
- Collecting opinions and attitudes from teachers and learners about the problem. A survey of teachers before they get the new L-design to use [157].
- Describes the full L-design for solving the problems, and how it is derived from the theory. It then reports on a test of the effectiveness of the full L-design [156].

A second kind of sequence in CER is captured in a single paper by Guzdial [51] (Chair's best paper award at ICER13). It reports a sequence of research studies on more or less the same course design, but progressing over ten years from (a) just describing the design, that it ran, and with "it seemed to go OK" judgements; to (b) formulating "hypotheses" or "design goals" about what its advantages might be (compared to other designs). This took the programme of work from reporting stories of what was done and experienced, to theories about specific effects and causes that were tested, leading to better theories and understanding. This was a ten year programme of research with direction and progress reported in a single paper, starting from a wholly practitioner experience viewpoint, and evolving into a researcher's viewpoint that established some useful generalisations.

3.11 Review papers.

Reviews are quite often published in CER: for instance the review of the use of theory in CER [90] which jointly won the ICER19 Chair's award for papers. In some other disciplines, major review papers can be places where theories are announced (by induction across many existing papers), or given a major relaunch. For example the journal *Psychological Review* [50] opens its self-description with the statement "Psychological Review publishes articles that make important theoretical contributions to any area of scientific psychology, including systematic evaluation of alternative theories".

Reviews may also be about L-designs rather than theories. One type is a review of evidence-based summaries of tested L-designs. These are important for promoting rational adoption without which there is little point to any applied research field. One example is [106] which presents a five page summary on "Computational Thinking in Introductory Physics", which besides being directed at practitioners who might adopt the cases it describes, also contributes to the theoretical concept of computational thinking in CER. Another kind of example is [40, 41] in the field of Physics

Education, about the design of teaching labs for students. These presentations are an overview with many examples of how it is now practicable to put on physics labs where the students have control over many aspects of it, rather than (as so often has been the case in the past) reproducing an exact set of instructions where the student is allowed no variation of what was done, and so no real connection with empirical work that answers research questions (which is actually the defining feature of the scientific method). In CER this would be the equivalent of "programming labs" where on the one hand students were given pre-written code and a fixed set of input values and asked to verify that the outputs were as predicted by the teacher; or on the other hand, given a test suite and program specification. A review in CER of "computing lab" activities might critically discuss this kind of issue, and indeed ask whether the students are being given the right kind of freedom of operation from an educational viewpoint.

Yet another kind of review might be useful in CER too. The journal “Brain and Behavioral Sciences” covers psychology, neuroscience, behavioral biology, cognitive science, artificial intelligence, linguistics, and philosophy (see [79] for an example). Instead of a systematic literature review, which guards against personal bias by stating the algorithm by which papers were included in the set reviewed, a different approach is taken. Each review’s authors argue a personal view of a field or issue plus arguments about what they think is important about it. What defends the paper against being a merely personal view is that multiple opinions and comments are invited and published with it at the same time (27 in the article cited as an example), followed by a response from the author(s) to the commentaries. This is public peer discussion collectively constituting a review of an issue, and carries its own authority based on more than the fact of publication which generally means approval by only two or three referees. This format might also perhaps be employed to address a general problem with "systematic reviews" which is that they seemingly cannot address what many readers would value: which are the few most valuable papers on a given topic?

Still other types of review paper might also have value. Since reviews are essentially reviews of the published literature, then in principle each type of citation found in papers might be considered as a distinct type of review topic. As mentioned in the Introduction (in §1.3), one type is the "Further reading" citation: essentially offering a pointer not to a single exemplar, nor to the first publication on a topic but to tutorial material for a reader who wants a useful introduction to a topic they know little of. While many research papers have no need of this type of reference, in a profoundly interdisciplinary field such as CER, it is fairly common to read a paper where you are expert in many of the things mentioned, but a stranger to some other aspect. If you have only used quantitative data but finally read a paper that used qualitative data to demonstrate something important to you, where can you find an introduction that is angled towards CER and not to the many other disciplines which use qualitative methods? If you work on a version of PI because that is the local tradition or requirement, then when you read of an alternative L-design in a way that makes you feel you should follow this up as a comparison to PI, where can you get an introduction to the alternative that is shorter than a book?

3.12 Tools to support learners or teachers.

Computer educational research is much more likely than other disciplinary areas to construct tools to help learners and assume that this is a contribution: after all, programming is a skill and a hobby for most CS academics. Guzdial and du Boulay argue that tools (along with methods and objectives) was one of the earliest if not the earliest theme in computing science education [52].

Some examples are:

- YAP3 1996: [167] The paper is about a plagiarism tool and is in the top 10 cited papers (with 185 citations) for SIGCSE.
- iSnap 2017: [127] The paper describes a tool that supports students with an Intelligent Tutoring system, and a pilot study of learner reactions. The paper received a Distinguished Paper Award at SIGCSE.
- PeerWise 2008: [27] The paper won a John Henry award at ICER08 for applying this in CS classes, and for presenting a tool to administer this . This paper described the software (PeerWise) to support this. It applied it in real courses and reported on that (an experience report). It also collected and reported quantitative data that supported the claim that learners using PeerWise engaged in deep learning, and performed better in exams, and so that there was a real educational gain because of it.
- Penmarked 2010: [123] The paper reports on the design, implementation, and repeated testing of a tool to help teachers mark their first year programming assignment. As discussed above under the contribution category "Requirements gathering for L-designs", it is notable for supporting teachers rather than learners, but also, in our opinion, for suggesting a generalisation to other disciplines with superficially very different types of assignments and marking.

4 USING THE SET OF CATEGORIES

We illustrate possible uses for our set of categories by three examples: two exercises and one analysis.

4.1 Imagine new bits of research for each category

Pick any category and then imagine a new piece of work that you don't know of anyone doing, which might be interesting and original and primarily in that one category. Then another one and repeat a few times.

4.2 Group exercise in using the categories to analyse a published or draft paper

Experience of reviewing papers, and sometimes of examining PhD theses, shows that researchers often do not realise the value of their work. No-one is surprised if it is less good than its author wishes, but sometimes it is good in a different way than they think. One way the category set could be used is to take a piece of work, hand it out to a set of CE researchers, and get them each to assign it to one or more categories. If they disagree then they may be about to invent a new category, but more often they will get practice at thinking about real contributions: identifying not the familiar surface elements in the written document (data, references to theory, ...) but the things of actual value and the nature of that value.

4.3 Using the categories to re-examine a published paper

A prize winning paper in ICER18 [101] argued that an over-emphasis on theory as what was valuable tended to suppress work on design itself. In terms of the 12 categories, that would translate as: too much *Theory for its own sake* (§3.1) tended to suppress *Research by building things* (§3.6).

The categories *Engineering for its own sake* (§3.5) and *CEPs* (§3.6) acknowledge that design and building designs are on an equal footing with theory, while *Unplanned observations* (§3.8) says more emphatically than the paper that some crucial contributions to CER come with no derivation from theory at all. On the other hand, if knowledge is to accumulate from these design activities, work on consolidating the experiences of construction as explicit *requirements* (§3.4) will surely be needed. With regard to *theory* (§3.1) we argued that CER is in effect interdisciplinary and so not locked into one type of theory, but on the contrary is continually using many different types from different disciplines.

The paper argued in effect that it is a zero sum game between theory and design, both expensive, which depends on the case rather than on a general rule. If it's a choice between creating large buildings or engineering theory then theory would be cheaper, but if it is creating variations on a user interface it is likely to be much faster to build, test, modify etc.

It may depend even more on what conclusions are to be drawn from the designs. If the context is improving a user interface with the hope of making it useful to everyone then each build and test will show a problem and the next version will be to solve that problem without making it worse for anyone, which can often be done with user interfaces, and done speedily. However it is not clear whether that would apply in CER since learners are often so different from each other. More scaffolding is good for a more ignorant learner but slows up a more knowledgeable one, so addressing this issue requires essentially personal tuition to keep adjusting the level of detail of the help given [169].

The category *Research programmes* (§3.10) suggests that combining the two approaches should be considered and might lead to a way to combine theory-focussed and application-focussed bits of research – a possibility not considered in the paper.

5 DISCUSSION: ASSESSING THE CONTRIBUTION OF THIS PAPER

5.1 The value of theory.

The original idea for this paper was to look at other kinds of research value than only theory. This means we have inevitably taken a position that looked for things other than theory, and demoted its prestige to just one of many types of research contribution. We wish to remind the reader nevertheless that, on the other hand, we have emphasised several times that theory is important, that in fact because CER is essentially an applied subject, theory has many different types of contribution: to understand the disciplinary content to be taught, but also it needs to use theories for each aspect that affects CE, and each aspect will need to use its own theory when available. In CER it may even be that exploring this richer than assumed role for theory will be the enduring value of this paper for some readers.

However because theory is often overvalued in our culture there are negative aspects to contend with. One is that because of this overvaluing, researchers tend to try to look as if they are involved in theory rather than actually being so. Malmi et al. [90] in its review of the contribution and use of theory in computing science education suggests that the vast majority of papers which cited theoretical constructs did so at a surface or cursory level with few discussing or considering them in any detail. This does not mean, however, that theory is not important even for the categories that most appear like its opposite. Louis Pasteur said that "in the fields of observation, chance only favours the mind which is prepared" [111]. That is, to notice what is interesting or what is unknown (the category of *unexpected observations* §3.8) depends upon a mind that is prepared by theoretical knowledge to recognise what is currently unexpected and contradicts a theory.

A more exact statement on the value of theory is that theory is good but only one element amongst a set of essential contributions. We should bear in mind that there are times, places and disciplines where theory has been over-valued [9], and others where it has been under-valued [102]. In the former case, large numbers of women died soon after childbirth even after direct proof by Semmelweis that hand-washing prevented this because direct evidence of improved prevention was not accepted in the absence of a proven new theory. While the authors of [9] give advice to readers on how to be more persuasive than Semmelweis, it is not only unethical but also fundamentally unscientific to give priority to theory over empirical observations. In the latter case, according to Nield ([102] p.130ff. especially pp.131-133 and 143-145), at one time American Geology academia would only accept a format in which new data were presented and then discussed in relation to multiple theories with no conclusion about which was correct. This was partly in reaction to European over-emphasis on theory, but it made it awkward to get the new breakthrough theory of continental drift published at all despite conclusive evidence.

Thus our aim should be to work to maintain a balance, and frequently review the state of our discipline. When too much attention is being paid to theory, we should seek to remedy that by focussing on other contributions: that was the original aim of this paper and expressed in the set of categories of contribution, of which theory and theoretical reasoning are only two of twelve. When we think too little attention is paid to theory, then we should promote it. Valentine [162], in the end, urged less mere practitioner reporting and more collection of data that could address (and prove or disprove) theoretical questions. A step beyond Valentine's call would be to urge work on new theories directly. However what might get researchers thinking about theories more sharply and immediately could be to propose experiments that test which of several contradictory predictions is true: an example of this was described in §3.2. Another way of stimulating attention to theory might be to ask specific theoretical questions about how a widely used L-design really works. Peer Instruction (PI) in CER has always combined it with a "flipped classroom" method, more properly called "Just In Time Teaching" (JITT) [104], where learners are required to read the relevant textbook content before the class. But surely we should look at the relative contributions to learning of each method. Crouch and Mazur [25] explicitly acknowledge and cite JITT, and furthermore say that the biggest additional increase in learning in PI came after they adopted a different technique for testing student preparation reading from JITT. A second question about how PI works in CER is about the design of the questions: in Mazur's PI in physics, the questions were based on a large prior research literature on misconceptions in physics, but that is not the case in CER. Does this matter? Is CER missing out on an important element of PI, or is it in fact not important? A third question is whether the peer discussion in PI is the only way to get this aspect of the design to work. Hunt's work [61] showed that considerable learning gains (reduction in learning time by up to 25%) could be created simply by having learners answer questions with not only their answer but additionally with an estimate of their certainty in its correctness. This might be a different method for creating that uncertainty about an idea which in PI is done by contrary opinions from peers.

5.2 Should we see categories as grouped in any way?

There is no single way to order the set of 12 categories. However there are several ways of grouping subsets. Some of these are each a kind of "axis" e.g.

- Research-Practice
- Theory-Observation
- Theory-Product

However these sound like oppositions whereas both in reality and in this paper, the collaboration and combination of categories stand out as more important.

Other kinds are clusters of categories by affinity:

- Theory and reasoning: *ACE* (§3.3), *Theory* (§3.1), *Predictions* (§3.2), *Reviews* (§3.11)
- Observations and data: *Planned observations* (§3.7), *Unplanned observations* (§3.8)
- Construction and design work: *CEPs* (§3.6), *Design work* (§3.5), *Requirements capture* (§3.4)
- Practitioners and practice i.e. user experience based reports: *Tools* (§3.12), *Practice papers* (§3.9)

The fact that these are all different ways of looking at the categories, which are all a bit interesting yet all different is perhaps an indication that our set does support breadth, does indicate the multi-factored nature of the set as a whole, and so indicates something of the breadth and variety of the terrain it attempts to represent. However thinking of these clusters may be echos of old ways of thinking rather than the new ways we hope will emerge.

5.3 Contributing named concepts

This paper was not the first, and neither will it be the last, in trying to understand the different types of contributions to knowledge in CER.

Kinnunen et al. [75] mentions a number of early CER categorisation systems but they seem to be looking backwards at what is, or are formed from categories of pre-existing types and not even from actual work in CER.

One of these was Valentine who invented a taxonomy of six types of paper [162]. His contribution has proved seminal in that several of the names he chose for the six categories, and their associated connotations, have been taken up and can be seen or heard today in referee reports and editorial meetings: “Marco Polo papers”, “John Henry” and “Philosophy”. This shows that discussing and recognising different types of contributions was already part of CER culture, and that it was augmented by adding three new terms.

So what is the value of our contribution in this paper? Has it moved on or is it yet another taxonomy paper, a "philosophy" paper, fueled by desk research and cemented in arguments from common experience? If even one of the proposed categories survives the test of time then that would at least be a contribution like Valentine’s to the CER community’s language, which is one of our aims.

However our set is not a taxonomy of types of paper observed but of contributions to knowledge. We would most like to shift discussion from types of paper, as if a paper could only make one contribution when several of our example papers fit in more than one category because they contain more than one contribution, and towards a focus on adding to knowledge, and discussing this explicitly.

5.4 Mapping the terrain of categories

The proposed set of categories is not a review of CER, neither is it a report on the recent direction of CER as a field, nor is it trying to steer what *ought* to be valued as a contribution in CER. The main aim is to map out a terrain of contributions with different landmarks, and prompt consideration of that terrain as a whole rather than narrowing down to preferred parts of it. There will, and perhaps should, be more such attempts at categorisation of contribution.

What is different in intention about the present paper is the attempt to create a space for people to decide for themselves what categories to choose to use, rather than try to fit in with categories for which there are existing precedents. We hope it is a virtue that many of the CER papers we mention in fact belong to more than one of our categories because it means our categories are closer to research functions performed than to actual fixed prescriptions for individual pieces of work (e.g. do an experiment and write it up), and this may turn out to be more useful. Essentially it is looking forward to what can exist rather than trying to fit only with what has gone before.

A published paper typically refers to multiple types of research contribution (e.g. past papers, experimental method, a new theoretical hypothesis) and often to multiple instances of one type. Many of our categories are instances of "research functions": each doing a job in the advancement of knowledge, and equally doing a job within a research project or a published paper. Our categories are a set of contrasting types of contribution to knowledge (the set being chosen to illustrate diversity, not to embody completeness).

6 CONCLUSION

This paper has focussed on proposing 12 categories of contributions to knowledge with the aim of stimulating thinking about this. If readers put down the paper and argue about the merits, defects, and utility of each one, then this aim will have been achieved. The hope is that this in turn will broaden ideas on how to contribute to CER, and will help the field by introducing ideas and terminology that might in turn help authors to be more explicit in their papers about how their work does (and does not) contribute.

If furthermore it emerges that people have tried the activities in the previous section and have found them productive, then that would be evidence of success for this paper's stated aims.

ACKNOWLEDGEMENTS

The authors would like to thank Allison Littlejohn and Maria Kallia for commenting on an early draft as well as Niall Barr, Quintin Cutts, Jack Parkinson, Lovisa Sundin, and Eric Yao for comments on later drafts. The authors are especially grateful to the anonymous referees and editors who gave extensive helpful suggestions for improvements.

REFERENCES

[1] Philip Adey. 2004. Evidence for Long-term Effects: Promises and Pitfalls. *Evaluation & research in education* 18, 1-2 (2004), 83–102. https://doi.org/10.1080/09500790408668310

[2] Alireza Ahadi and Raymond Lister. 2013. Geek genes, prior knowledge, stumbling points and learning edge momentum: parts of the one elephant?. In *Proceedings of the ninth annual international ACM conference on International computing education research*. Association for Computing Machinery, New York, NY, USA, 123–128. https://doi.org/10.1145/2493394.2493416

[3] Deborah J Armstrong and Bill C Hardgrave. 2007. Understanding mindshift learning: the transition to object-oriented development. *MIS Quarterly* 31, 3 (2007), 453–474. https://www.jstor.org/stable/25148803

[4] Elliot Aronson et al. 1978. *The jigsaw classroom.* Sage, California, CA, USA. https://www.jigsaw.org/

[5] Yirsaw Ayalew, Ethel Tshukudu, and Moemedi Lefoanea. 2018. Factors affecting programming performance of first year students at a University in Botswana. *African Journal of Research in Mathematics, Science and Technology Education* 22, 3 (2018), 363–373. https://doi.org/10.1080/18117295.2018.1540169

[6] Ashok R Basawapatna, Alexander Repenning, Kyu Han Koh, and Hilarie Nickerson. 2013. The zones of proximal flow: guiding students through a space of computational thinking skills and challenges. In *Proceedings of the ninth annual international ACM conference on International Computing Education Research*. Association for Computing Machinery, New York, NY, USA, 67–74.

[7] Mordechai Ben-Ari. 2001. Constructivism in computer science education. *Journal of Computers in Mathematics and Science Teaching* 20, 1 (2001), 45–73.

[8] Mordechai Ben-Ari. 2004. Situated learning in computer science education. *Computer Science Education* 14, 2 (2004), 85–100.

[9] M Best and D Neuhauser. 2004. *Ignaz Semmelweis and the birth of infection control*. BMJ. Retrieved 5 Jan, 2022 from https://qualitysafety.bmj.com/content/13/3/233

[10] K Yusuf Billah and Robert H Scanlan. 1991. Resonance, Tacoma Narrows bridge failure, and undergraduate physics textbooks. *American Journal of Physics* 59, 2 (1991), 118–124.

[11] B.S. Bloom. 1984. The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational researcher* 13, 6 (1984), 4–16. https://doi.org/10.2307/1175554

[12] BMJ. 2011. *Is The BMJ the right journal for my research article?* BMJ. Retrieved June 10, 2021 from https://www.bmj.com/about-bmj/resources-authors/bmj-right-journal-my-research-article

[13] Jonas Boustedt, Robert McCartney, Katherine Deibel, Jim Huggins, Beth Simon, Suzanne Westbrook, and The Mystery Presenter. 2009. It seemed like a good idea at the time. *ACM SIGCSE Bulletin* 41, 1 (2009), 265–266.

[14] Jonas Boustedt, Robert McCartney, Josh Tenenberg, Scott D Anderson, Caroline M Eastman, Daniel D Garcia, Paul V Gestwicki, and Margaret S Menzin. 2008. It seemed like a good idea at the time. In *Proceedings of the 39th SIGCSE technical symposium on Computer science education*. ACM New York, NY, USA, New York, NY, USA, 528–529.

[15] Jonas Boustedt, Robert McCartney, Josh Tenenberg, Titus Winters, Stephen Edwards, Briana B Morrison, David R Musicant, Ian Utting, and Carol Zander. 2007. It seemed like a good idea at the time. *ACM SIGCSE Bulletin* 39, 1 (2007), 346–347.

[16] Steven G Brandl. 2017. *Criminal investigation.* SAGE Publications, California, CA, USA.

[17] Ann L. Brown. 1992. Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences* 2, 2 (1992), 141–178. https://doi.org/10.1207/s15327809jls0202_2

[18] Neil CC Brown, Amjad Altadmri, Sue Sentance, and Michael Kölling. 2018. Blackbox, five years on: An evaluation of a large-scale programming data collection project. In *Proceedings of the 2018 ACM Conference on International Computing Education Research*. Association for Computing Machinery, New York, NY, USA, 196–204.

[19] John Carroll and Mary Beth Rosson. 1992. Getting Around the Task-Artifact Cycle: How to Make Claims and Design by Scenario. *ACM Trans. Inf. Syst.* 10 (04 1992), 181–212. https://doi.org/10.1145/146802.146834

[20] Michelene T. H. Chi and Ruth Wylie. 2014. The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational psychologist* 49, 4 (2014), 219–243. https://doi.org/10.1080/00461520.2014.965823

[21] Matthew M. Chingos and Paul E. Peterson. 2011. It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of education review* 30, 3 (2011), 449–465. https://doi.org/10.1016/j.econedurev.2010.12.010

[22] Mike Clancy, John Stasko, Mark Guzdial, Sally Fincher, and Nell Dale. 2001. Models and areas for CS education research. *Computer Science Education* 11, 4 (2001), 323–341. https://www.tandfonline.com/doi/pdf/10.1076/csed.11.4.323.3827

[23] Mike Cole and Sylvia Scribner. 1977. Cross-cultural studies of memory and cognition. In *Perspectives on the development of memory and cognition*. Erlbaum: Hillsdale, N.J., Aylesbury, Bucks, UK, Chapter 8, 239–271.

[24] Betty Collis and JCMM Moonen. 2005. *An on-going journey: Technology as a learning workbench*. University of Twente. Retrieved June 10, 2021 from https://research.utwente.nl/en/publications/an-on-going-journey-technology-as-a-learning-workbench

[25] Catherine H. Crouch and Eric Mazur. 2001. Peer Instruction: Ten years of experience and results. *American journal of physics* 69, 9 (2001), 970–977. https://doi.org/10.1119/1.1374249

[26] Quintin Cutts, Matthew Barr, Mireilla Bikanga Ada, Peter Donaldson, Steve Draper, Jack Parkinson, Jeremy Singer, and Lovisa Sundin. 2019. Experience report: thinkathon - countering an í got it workingḿentality with pencil-and-paper exercises. *ACM Inroads* 10, 4 (2019), 66–73.

[27] Paul Denny, John Hamer, Andrew Luxton-Reilly, and Helen Purchase. 2008. PeerWise: Students Sharing Their Multiple Choice Questions. In *Proceedings of the Fourth International Workshop on Computing Education Research* (Sydney, Australia) *(ICER '08)*. Association for Computing Machinery, New York, NY, USA, 51–58. https://doi.org/10.1145/1404520.1404526

[28] Gordon R. Dickson. 1971. *Tactics of Mistake*. Sphere Books, Aylesbury, Bucks, UK, Chapter 1–2, 11–14. Cletus is the protagonist, and debates theory vs. practical knowledge with others.

[29] Edsger W Dijkstra et al. 1989. On the cruelty of really teaching computing science. *Commun. ACM* 32, 12 (1989), 1398–1404.

[30] Steve Draper. 1988. What's going on in everyday explanation? In *Analysing everyday explanation*. SAGE Publications, California, CA, USA, Chapter 2, 15–31.

[31] Steve Draper. 1994. Constructivism, other theories of the teaching and learning process, and their relationships. https://www.psy.gla.ac.uk/~steve/courses/apechos/NATO.ho20.pdf

[32] Steve Draper. ?1994. funding. Nicol&Draper? Where are my copies of Nicol papers. https://www.psy.gla.ac.uk/~steve/courses/apechos/NATO.ho20.pdf

[33] Steve Draper. 2009. What are learners actually regulating when given feedback? *British Journal of Educational Technology* 40, 2 (2009), 306–315. https://doi.org/10.1111/j.1467-8535.2008.00930.x

[34] Steve Draper. 2020. Jigsaw classroom technique, within Some Local Educational Research. https://www.psy.gla.ac.uk/~steve/localed/#jigsaw

[35] Stephen W. Draper. 2009. Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology* 40, 2 (2009), 285–293. https://doi.org/10.1111/j.1467-8535.2008.00920.x

[36] Underwood Dudley. 2002. The world's first mathematics textbook. *Math Horizons* 9, 4 (2002), 8–11.

[37] Durkheim_(Wikipedia). 2021. Suicide (Durkheim book). https://en.wikipedia.org/wiki/Suicide_(Durkheim_book)

[38] Sebastian Dziallas and Sally Fincher. 2016. Aspects of Graduateness in Computing Students' Narratives. In *Proceedings of the 2016 ACM Conference on International Computing Education Research* (Melbourne, VIC, Australia) *(ICER '16)*. Association for Computing Machinery, New York, NY, USA, 181–190. https://doi.org/10.1145/2960310.2960317

[39] Anna Eckerdal, Michael Thuné, and Anders Berglund. 2005. What does it take to learn 'programming thinking'?. In *Proceedings of the first international workshop on Computing education research*. Association for Computing Machinery, New York, NY, USA, 135–142.

[40] Eugenia Etkina. 2015. Eugenia Etkina Millikan Lecture 2014: Students of physics—Listeners, observers, or collaborative participants? https://www.youtube.com/watch?v=gn1M23gifjU

[41] Eugenia Etkina. 2015. Millikan award lecture: Students of physics—Listeners, observers, or collaborative participants in physics scientific practices? *American Journal of Physics* 83, 4 (2015), 669–679. https://doi.org/10.1119/1.4923432

[42] A. C. Fabian. 2009. Serendipity in Astronomy. arXiv:0908.2784 [physics.pop-ph]

[43] Bernard J Feldman. 2003. What to say about the Tacoma Narrows Bridge to your introductory physics class. *The Physics Teacher* 41, 2 (2003), 92–96.

[44] Peter J Fensham. 2004. *Defining an identity: The evolution of science education as a field of research*. Vol. 20. Springer Science & Business Media, New York, NY, USA.

[45] Sally Fincher and Marian Petre. 2004. *Computer science education research*. CRC Press, Boca Raton, Florida, USA.

[46] Sally Fincher and Josh Tenenberg. 2006. Using theory to inform capacity-building: Bootstrapping communities of practice in computer science education research. *Journal of Engineering Education* 95, 4 (2006), 265–277.

[47] Sally A Fincher and Anthony Robins. 2019. An important and timely field. In *The Cambridge Handbook of Computing Education Research, Sally A. Fincher and Anthony V. Robins (Eds.). Cambridge University Press, Cambridge, Chapter 0*. Cambridge University Press, Cambridge, England, UK.

[48] Sally A Fincher and Anthony V Robins. 2019. An Important and Timely Field. In *The Cambridge handbook of computing education research*. Cambridge University Press, London, UK, Chapter 0, 1–8.
[49] Peter Galison et al. 1997. *Image and logic: A material culture of microphysics*. University of Chicago Press, Chicago, IL, USA.
[50] Elena L. Grigorenko. 1984. Psychological Review. https://www.apa.org/pubs/journals/rev
[51] Mark Guzdial. 2013. Exploring hypotheses about media computation. In *Proceedings of the ninth annual international ACM conference on International computing education research (ICER '13)*. Association for Computing Machinery, New York, NY, USA, 19–26. https://doi.org/10.1145/2493394.2493397
[52] Mark Guzdial and Benedict du Boulay. 2019. The history of computing education research. *The Cambridge handbook of computing education research* 2019 (2019), 11–39.
[53] Richard R Hake. 1998. Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics* 66, 1 (1998), 65.
[54] John Hamer, Quintin Cutts, Jana Jackova, Andrew Luxton-Reilly, Robert McCartney, Helen Purchase, Charles Riedesel, Mara Saeli, Kate Sanders, and Judithe Sheard. 2008. Contributing Student Pedagogy. *SIGCSE Bull.* 40, 4 (Nov. 2008), 194–212. https://doi.org/10.1145/1473195.1473242
[55] John Hamer, Judithe Sheard, Helen Purchase, and Andrew Luxton-Reilly. 2012. Contributing student pedagogy. *Computer Science Education* 22, 4 (2012), 315–318. https://doi.org/10.1080/08993408.2012.727709
[56] Eric A Hanushek. 2011. The economic value of higher teacher quality. *Economics of Education review* 30, 3 (2011), 466–479.
[57] Qiang Hao, David H Smith IV, Naitra Iriumi, Michail Tsikerdekis, and Andrew J Ko. 2019. A systematic investigation of replications in computing education research. *ACM Transactions on Computing Education (TOCE)* 19, 4 (2019), 1–18.
[58] Stephen P Heyneman, Joseph P Farrell, and Manuel A Sepulveda-Stuardo. 1981. Textbooks and achievement in developing countries: What we know. *Journal of curriculum studies* 13, 3 (1981), 227–246.
[59] Karl Hostetler. 2005. What is "good" education research? *Educational researcher* 34, 6 (2005), 16–21. https://journals.sagepub.com/doi/pdf/10.3102/0013189X034006016
[60] Geoffrey Howson. 2013. The development of mathematics textbooks: historical reflections from a personal perspective. *ZDM* 45, 5 (2013), 647–658.
[61] Darwin P Hunt. 1982. Effects of human self-assessment responding on learning. *Journal of Applied Psychology* 67, 1 (1982), 75. https://doi.org/10.1037/0021-9010.67.1.75
[62] William H Ittelson. 1968. *The Ames demonstrations in perception; Together With an Interpretative Manual*. Princeton University Press, Princeton, NJ, USA.
[63] William H Ittelson and Franklin P Kilpatrick. 1951. Experiments in perception. *Scientific American* 185, 2 (1951), 50–56. https://www.jstor.org/stable/24945240
[64] N Jiang. 2000. Lexical representation and development in a second language. *Applied Linguistics* 21, 1 (03 2000), 47–77. https://doi.org/10.1093/applin/21.1.47 arXiv:https://academic.oup.com/applij/article-pdf/21/1/47/437225/210047.pdf
[65] Nan Jiang. 2002. Form–meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition* 24, 4 (2002), 617–637. https://doi.org/10.1017/S0272263102004047
[66] Nan Jiang. 2004. Semantic transfer and its implications for vocabulary teaching in a second language. *The modern language journal* 88, 3 (2004), 416–432.
[67] Jeff Johnson and Austin Henderson. 1993. Conceptual models: begin by designing what to design. *interactions* 9, 1 (1993), 25–32. https://doi.org/10.1145/503355.503366
[68] Yasmin Kafai, Chris Proctor, and Debora Lui. 2020. From theory bias to theory dialogue: embracing cognitive, situated, and critical framings of computational thinking in K-12 CS education. *ACM Inroads* 11, 1 (2020), 44–53.
[69] Yasmin B Kafai, David DeLiema, Deborah A Fields, Gary Lewandowski, and Colleen Lewis. 2019. Rethinking debugging as productive failure for CS education. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 169–170.
[70] Maria Kallia and Quintin Cutts. 2021. Re-Examining Inequalities in Computer Science Participation from a Bourdieusian Sociological Perspective. In *Proceedings of the 14th ACM Conference on International Computing Education Research (ICER 2021)* (Glasgow, UK) *(ICER '21)*. Association for Computing Machinery, New York, NY, USA, 14 pages. https://doi.org/10.1145/3446871.3469763
[71] Maria Kallia and Quintin Cutts. 2021. Re-Examining Inequalities in Computer Science Participation from a Bourdieusian Sociological Perspective. In *Proceedings of the 17th ACM Conference on International Computing Education Research* (Virtual Event, USA) *(ICER 2021)*. Association for Computing Machinery, New York, NY, USA, 379–392. https://doi.org/10.1145/3446871.3469763
[72] John G Kappenman, Lawrence J Zanetti, and William A Radasky. 1997. Geomagnetic storms can threaten electric power grid. *Earth in Space* 9, 7 (1997), 9–11.
[73] Manu Kapur. 2008. Productive failure. *Cognition and instruction* 26, 3 (2008), 379–424.
[74] Jeffrey D. Karpicke and Janell R. Blunt. 2011. Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science* 331, 6018 (2011), 772–775. https://doi.org/10.1126/science.1199327 arXiv:https://science.sciencemag.org/content/331/6018/772.full.pdf
[75] Päivi Kinnunen, Veijo Meisalo, and Lauri Malmi. 2010. Have We Missed Something? Identifying Missing Types of Research in Computing Education. In *Proceedings of the Sixth International Workshop on Computing Education Research* (Aarhus, Denmark) *(ICER '10)*. Association for Computing Machinery, New York, NY, USA, 13–22. https://doi.org/10.1145/1839594.1839598
[76] Donald Ervin Knuth. 1997. *The art of computer programming*. Vol. 3. Pearson Education, New York, NY, USA.

[77] Donald E Knuth. 2007. Computer programming as an art. In *ACM Turing award lectures*. Association for Computing Machinery, New York, NY, USA, 1974.

[78] Thomas Kuhn. 2021. *The structure of scientific revolutions*. Princeton University Press, Princeton, NJ, USA.

[79] Brenden M. Lake, Tomer D. Ullman, Joshua B. Tenenbaum, and Samuel J. Gershman. 2017. Building machines that learn and think like people. *Behavioral and Brain Sciences* 40 (2017), e253. https://doi.org/10.1017/S0140525X16001837

[80] Cynthia Bailey Lee, Saturnino Garcia, and Leo Porter. 2013. Can peer instruction be effective in upper-division computer science courses? *ACM Transactions on Computing Education (TOCE)* 13, 3 (2013), 1–22. https://doi.org/10.1145/2499947.2499949

[81] Dalit Levy. 2001. Insights and conflicts in discussing recursion: A case study. *Computer Science Education* 11, 4 (2001), 305–322.

[82] Colleen M Lewis, Tim Bell, Paulo Blikstein, Adam S Carter, Katrina Falkner, Sally A Fincher, Kathi Fisler, Mark Guzdial, Patricia Haden, Sepehr Hejazi Moghadam, et al. 2020. The Cambridge Handbook of Computing Education Research Summarized in 75 minutes. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. Association for Computing Machinery, New York, NY, USA, 323–324.

[83] Soohyun Nam Liao, William G. Griswold, and Leo Porter. 2018. Classroom Experience Report on Jigsaw Learning. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education* (Larnaca, Cyprus) *(ITiCSE 2018)*. Association for Computing Machinery, New York, NY, USA, 302–307. https://doi.org/10.1145/3197091.3197118

[84] Raymond Lister, Beth Simon, Errol Thompson, Jacqueline L. Whalley, and Christine Prasad. 2006. Not Seeing the Forest for the Trees: Novice Programmers and the SOLO Taxonomy. *SIGCSE Bull.* 38, 3 (June 2006), 118–122. https://doi.org/10.1145/1140123.1140157

[85] Kim Man Lui and Keith C.C. Chan. 2006. Pair programming productivity: Novice–novice vs. expert–expert. *International Journal of Human-Computer Studies* 64, 9 (2006), 915–925. https://doi.org/10.1016/j.ijhcs.2006.04.010

[86] Joseph Maguire and Steve Draper. 2022. Promoting Engagement in Remote Computing Ethics Education. In *Proceedings of the 3rd Conference on Computing Education Practice* (Durham, United Kingdom) *(CEP 2022)*. Association for Computing Machinery, New York, NY, USA, 4 pages.

[87] Joseph Maguire, Steve Draper, and Quintin Cutts. 2019. What Do We Do When We Teach Software Engineering?. In *Proceedings of the 1st UK & Ireland Computing Education Research Conference*. ACM, New York, NY, USA, 1–7.

[88] Joseph Maguire, Rosanne English, and Steve Draper. 2019. Data protection and privacy regulations as an inter-active-constructive practice. In *Proceedings of the 3rd Conference on Computing Education Practice*. ACM, New York, NY, USA, 1–4.

[89] Lauri Malmi, Judy Sheard, Roman Bednarik, Juha Helminen, Päivi Kinnunen, Ari Korhonen, Niko Myller, Juha Sorva, and Ahmad Taherkhani. 2014. Theoretical underpinnings of computing education research: what is the evidence?. In *Proceedings of the tenth annual conference on International computing education research*. Association for Computing Machinery, New York, NY, USA, 27–34.

[90] Lauri Malmi, Judy Sheard, Päivi Kinnunen, and Jane Sinclair. 2019. Computing Education Theories: What Are They and How Are They Used?. In *Proceedings of the 2019 ACM Conference on International Computing Education Research*. Association for Computing Machinery, New York, NY, USA, 187–197.

[91] Lauri Malmi, Judy Sheard, Päivi Kinnunen, and Jane Sinclair. 2020. Theories and models of emotions, attitudes, and self-efficacy in the context of programming education. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*. Association for Computing Machinery, New York, NY, USA, 36–47.

[92] Lauren E. Margulieux. 2019. Spatial Encoding Strategy Theory: The Relationship between Spatial Skill and STEM Achievement. In *Proceedings of the 2019 ACM Conference on International Computing Education Research* (Toronto ON, Canada) *(ICER '19)*. Association for Computing Machinery, New York, NY, USA, 81–90. https://doi.org/10.1145/3291279.3339414

[93] Eric Mazur. 1997. *Peer Instruction: A User's Manual*. Prentice Hall, Upper Saddle River, NJ, USA.

[94] Afaf I Meleis. 1992. Directions for nursing theory development in the 21st century. *Nursing Science Quarterly* 5, 3 (1992), 112–117.

[95] Afaf Ibrahim Meleis. 2011. *Theoretical nursing: Development and progress*. Lippincott Williams & Wilkins, Wolters Kluwer, Alphen aan den Rijn, Netherlands.

[96] Alejandra Mendez, Ethel Tshukudu, Carolina Moreira, Weena Naowaprateep, Alan Peterfreund, and Brianna Johnston. 2021. *Identifying Opportunities and Potential Roadblocks for CSEd Professionals*. Association for Computing Machinery, New York, NY, USA, 433–434. https://doi.org/10.1145/3446871.3469789

[97] Leon S Moisseiff and Frederick Lienhard. 1933. Suspension bridges under the action of lateral forces. *Transactions of the American Society of Civil Engineers* 98, 2 (1933), 1080–1095.

[98] Charles R Morris. 2012. *The dawn of innovation: The first American industrial revolution*. PublicAffairs, New York, NY, USA.

[99] MRC. 2000. Framework for development and evaluation of RCTs for complex interventions to improve health. https://www.mrc.ac.uk/documents/pdf/rcts-for-complex-interventions-to-improve-health/

[100] Thomas L Naps, Guido Rößling, Vicki Almstrum, Wanda Dann, Rudolf Fleischer, Chris Hundhausen, Ari Korhonen, Lauri Malmi, Myles McNally, Susan Rodger, et al. 2002. Exploring the role of visualization and engagement in computer science education. In *Working group reports from ITiCSE on Innovation and technology in computer science education*. ACM, New York, NY, USA, 131–152.

[101] Greg L Nelson and Andrew J Ko. 2018. On use of theory in computing education research. In *Proceedings of the 2018 ACM Conference on International Computing Education Research*. Association for Computing Machinery, New York, NY, USA, 31–39.

[102] Ted Nield. 2007. *Supercontinent: Ten billion years in the life of our planet*. Granta, Croydon, UK.

[103] Ray P. Norris. 2017. Discovering the Unexpected in Astronomical Survey Data. *Publications of the Astronomical Society of Australia* 34 (2017), e007. https://doi.org/10.1017/pasa.2016.63

[104] Gregor M Novak, Evelyn T Patterson, Andrew D Gavrin, and Wolfgang Christian. 1999. *Just-in-Time Teaching: Blending Active Learning and Web Technology.* Prentice-Hall, Upper Saddle River, NJ. http://webphysics.iupui.edu/jitt/jitt.html

[105] Éadaoin O'Brien, Niamh Nic Daeid, and Sue Black. 2015. Science in the court: pitfalls, challenges and solutions. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370, 1674 (2015), 20150062.

[106] C.M. Orban and R.M. Teeling-Smith. 2009. Computational Thinking in Introductory Physics. *The Physics Teacher* 58 (2009), 247–251. https://doi.org/10.1119/1.5145470

[107] Zacharoula Papamitsiou, Michail Giannakos, Simon, and Andrew Luxton-Reilly. 2020. Computing Education Research Landscape through an Analysis of Keywords. In *Proceedings of the 2020 ACM Conference on International Computing Education Research.* ACM, New York, NY, USA, 102–112.

[108] Seymour Papert. 1980. *Mindstorms: Children, Computers, And Powerful Ideas.* Basic Books, New York, NY, USA.

[109] Eugene N. Parker. 1958. Dynamics of the Interplanetary Gas and Magnetic Fields. *The Astrophysics Journal* 128 (Nov. 1958), 664–676. https://adsabs.harvard.edu/pdf/1958apj...128..664p

[110] E. N. Parker. 1997. The martial art of scientific publication. *Eos, Transactions American Geophysical Union* 78, 37 (1997), 391–395. https://doi.org/10.1029/97EO00251

[111] L Pasteur. 1939. Speech delivered at Douai on December 7, 1854 on the occasion of his formal inauguration to the Faculty of Letters of Douai and the Faculty of Sciences of Lille.

[112] Elizabeth Patitsas, Jesse Berlin, Michelle Craig, and Steve Easterbrook. 2016. Evidence That Computer Science Grades Are Not Bimodal. In *Proceedings of the 2016 ACM Conference on International Computing Education Research* (Melbourne, VIC, Australia) *(ICER '16).* Association for Computing Machinery, New York, NY, USA, 113–121. https://doi.org/10.1145/2960310.2960312

[113] Judea Pearl. 2000. *Causality: Models, Reasoning and Inference.* Cambridge University Press, Cambridge, England, UK.

[114] Judea Pearl. 2009. Causal inference in statistics: An overview. *Statistics Surveys* 3 (2009), 96–146. https://doi.org/10.1214/09-SS057

[115] Judea Pearl. 2011. A calculus of probabilitistic and causal reasoning. https://amturing.acm.org/award_winners/pearl_2658896.cfm Turing award lecture.

[116] Judea Pearl and Dana Mackenzie. 2018. *The book of why: The new science of cause and effect.* Basic Books, New York, NY, USA.

[117] Arnold Pears, Stephen Seidman, Crystal Eney, Päivi Kinnunen, and Lauri Malmi. 2005. Constructing a core literature for computing education research. *ACM SIGCSE Bulletin* 37, 4 (2005), 152–161. https://dl.acm.org/doi/pdf/10.1145/1113847.1113893

[118] Katherine K Perkins and Carl E Wieman. 2005. The surprising impact of seat location on student performance. *The Physics Teacher* 43, 1 (2005), 30–33. https://doi.org/10.1119/1.1845987

[119] Marian Petre, Kate Sanders, Robert McCartney, Marzieh Ahmadzadeh, Cornelia Connolly, Sally Hamouda, Brian Harrington, Jérémie Lumbroso, Joseph Maguire, Lauri Malmi, et al. 2020. Mapping the Landscape of Peer Review in Computing Education Research. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education.* ACM, New York, NY, USA, 173–209.

[120] Henry Petroski. 1985. *To engineer is human: The role of failure in successful design.* Vintage, London, England, UK.

[121] Henry Petroski. 2009. Engineering: Tacoma Narrows Bridges. *American Scientist* 97, 2 (2009), 103–107.

[122] Robert M. Pirsig. 1974. *Zen and the art of motorcycle maintenance.* William Morrow and Company, New York, NY, USA.

[123] Beryl Plimmer. 2010. A Comparative Evaluation of Annotation Software for Grading Programming Assignments. In *Proceedings of the Eleventh Australasian Conference on User Interface - Volume 106* (Brisbane, Australia) *(AUIC '10).* Australian Computer Society, Inc., AUS, 14–22.

[124] Beryl Plimmer and Paul Mason. 2006. A Pen-Based Paperless Environment for Annotating and Marking Student Assignments. In *Proceedings of the 7th Australasian User Interface Conference - Volume 50* (Hobart, Australia) *(AUIC '06).* Australian Computer Society, Inc., AUS, 37–44.

[125] Karl Popper. 2005. *The logic of scientific discovery.* Routledge, London, England, UK.

[126] Leo Porter, Dennis Bouvier, Quintin Cutts, Scott Grissom, Cynthia Lee, Robert McCartney, Daniel Zingaro, and Beth Simon. 2016. A multi-institutional study of peer instruction in introductory computing. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education.* Association for Computing Machinery, New York, NY, USA, 358–363.

[127] Thomas W. Price, Yihuan Dong, and Dragan Lipovac. 2017. ISnap: Towards Intelligent Tutoring in Novice Programming Environments. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (Seattle, Washington, USA) *(SIGCSE '17).* Association for Computing Machinery, New York, NY, USA, 483–488. https://doi.org/10.1145/3017680.3017762

[128] T.S. Rappaport, A. Annamalai, R.M. Buehrer, and W.H. Tranter. 2002. Wireless communications: past events and a future perspective. *IEEE Communications Magazine* 40, 5 (2002), 148–161. https://doi.org/10.1109/MCOM.2002.1006984

[129] Scott Reeves, Mathieu Albert, Ayelet Kuper, and Brian David Hodges. 2008. Why use theories in qualitative research?

[130] Barbara J Reys, Robert E Reys, and Oscar Chavez. 2004. Why Mathematics Textbooks Matter. *Educational Leadership* 61, 5 (2004), 61–66.

[131] Anthony Robins. 2010. Learning edge momentum: A new account of outcomes in CS1. *Computer Science Education* 20, 1 (2010), 37–71.

[132] Bertrand Russell. 1914. *Our Knowledge of the External World as a Field for Scientific Method in Philosophy.* Open Court Publishing, Chicago and London.

[133] Kate Sanders, Jonas Boustedt, Anna Eckerdal, Robert McCartney, and Carol Zander. 2017. Folk pedagogy: Nobody doesn't like active learning. In *Proceedings of the 2017 ACM Conference on International Computing Education Research.* ACM, New York, NY, USA, 145–154.

[134] Kate Sanders, Jonas Boustedt, Anna Eckerdal, Robert McCartney, and Carol Zander. 2017. Folk Pedagogy: Nobody Doesn't Like Active Learning. In *Proceedings of the 2017 ACM Conference on International Computing Education Research* (Tacoma, Washington, USA) *(ICER '17).* Association for

Computing Machinery, New York, NY, USA, 145–154. https://doi.org/10.1145/3105726.3106192

[135] Adolf J Schwab and Peter Fischer. 1998. Maxwell, Hertz, and German radio-wave history. *Proc. IEEE* 86, 7 (1998), 1312–1318.

[136] Richard Scott. 2001. *In the wake of Tacoma: suspension bridges and the quest for aerodynamic stability.* American Society of Civil Engineers, New York, NY, USA.

[137] Anna Sfard. 1998. On two metaphors for learning and the dangers of choosing just one. *Educational researcher* 27, 2 (1998), 4–13. https://doi.org/10.3102/0013189X027002004

[138] Dudley Shapere. 1964. The structure of scientific revolutions. *The Philosophical Review* 73, 3 (1964), 383–394.

[139] Michael Shayer. 2016. Problems and issues in intervention studies. In *Neo-Piagetian theories of cognitive development: Implications and applications for education.* Routledge, London, England, UK, Chapter 6, 107–121.

[140] Judy Sheard. 2020. Twenty-Four Years of ITiCSE Papers. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education.* Association for Computing Machinery, New York, NY, USA, 5–11.

[141] Bruce L Sherin. 2001. A comparison of programming languages and algebraic notation as expressive languages for physics. *International Journal of Computers for Mathematical Learning* 6, 1 (2001), 1–61.

[142] ACM SIGCSE. 2005. *ITiCSE announces awards first 25 years.* ACM SIGCSE. Retrieved June 10, 2021 from https://sigcse.org/about/news/Top-iticse-papers.html

[143] Simon, Angela Carbone, Michael de Raadt, Raymond Lister, Margaret Hamilton, and Judy Sheard. 2008. Classifying Computing Education Papers: Process and Results. In *Proceedings of the Fourth International Workshop on Computing Education Research* (Sydney, Australia) *(ICER '08).* Association for Computing Machinery, New York, NY, USA, 161–172. https://doi.org/10.1145/1404520.1404536

[144] M. K. Smith, W. B. Wood, W. K. Adams, C. Wieman, J. K. Knight, N. Guild, and T. T. Su. 2009. Why Peer Discussion Improves Student Performance on In-Class Concept Questions. *Science (American Association for the Advancement of Science)* 323, 5910 (2009), 122–124. https://doi.org/10.1126/science.1165919

[145] J Srinivasan. 2001. Sadi Carnot and the second law of thermodynamics. *Resonance* 6, 11 (2001), 42–48.

[146] Bharath Sriraman and Lyn English. 2009. *Theories of mathematics education: Seeking new frontiers.* Springer Science & Business Media, New York, NY, USA.

[147] Donald E Stokes. 2011. *Pasteurś quadrant: Basic science and technological innovation.* Brookings Institution Press, Washington, DC, USA.

[148] Patrick Suppes. 1974. The place of theory in educational research. *Educational researcher* 3, 6 (1974), 3–10.

[149] Claudia Szabo, Nickolas Falkner, Andrew Petersen, Heather Bort, Kathryn Cunningham, Peter Donaldson, Arto Hellas, James Robinson, and Judy Sheard. 2019. Review and use of learning theories within computer science education research: primer for researchers and practitioners. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education.* Association for Computing Machinery, New York, NY, USA, 89–109. https://dl.acm.org/doi/10.1145/3344429.3372504

[150] Josh Tenenberg and Yifat Ben-David Kolikant. 2014. Computer Programs, Dialogicality, and Intentionality. In *Proceedings of the Tenth Annual Conference on International Computing Education Research* (Glasgow, Scotland, United Kingdom) *(ICER '14).* Association for Computing Machinery, New York, NY, USA, 99–106. https://doi.org/10.1145/2632320.2632351

[151] Vincent Tinto. 1975. Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research* 45, 1 (1975), 89–125. https://doi.org/10.3102/00346543045001089

[152] Leslie Tomory. 2013. Fostering a new industry in the Industrial Revolution: Boulton & Watt and gaslight 1800–1812. *The British Journal for the History of Science* 46, 2 (2013), 199–229.

[153] Ethel Tshukudu. 2019. Towards a Model of Conceptual Transfer for Students Learning New Programming Languages. In *Proceedings of the 2019 ACM Conference on International Computing Education Research* (Toronto ON, Canada) *(ICER '19).* Association for Computing Machinery, New York, NY, USA, 355–356. https://doi.org/10.1145/3291279.3339437

[154] Ethel Tshukudu and Quintin Cutts. 2020. Semantic Transfer in Programming Languages: Exploratory Study of Relative Novices. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education* (Trondheim, Norway) *(ITiCSE '20).* Association for Computing Machinery, New York, NY, USA, 307–313. https://doi.org/10.1145/3341525.3387406

[155] Ethel Tshukudu and Quintin Cutts. 2020. Understanding Conceptual Transfer for Students Learning New Programming Languages. In *Proceedings of the 2020 ACM Conference on International Computing Education Research* (Virtual Event, New Zealand) *(ICER '20).* Association for Computing Machinery, New York, NY, USA, 227–237. https://doi.org/10.1145/3372782.3406270

[156] Ethel Tshukudu, Quintin Cutts, and Mary Ellen Foster. 2021. Evaluating a Pedagogy for Improving Conceptual Transfer and Understanding in a Second Programming Language Learning Context. In *21st Koli Calling International Conference on Computing Education Research* (Joensuu, Finland) *(Koli Calling '21).* Association for Computing Machinery, New York, NY, USA, Article 26, 10 pages. https://doi.org/10.1145/3488042.3488050

[157] Ethel Tshukudu, Quintin Cutts, Olivier Goletti, Alaaeddin Swidan, and Felienne Hermans. 2021. *Teachers' Views and Experiences on Teaching Second and Subsequent Programming Languages.* Association for Computing Machinery, New York, NY, USA, 294–305. https://doi.org/10.1145/3446871.3469752

[158] Ethel Tshukudu and Siri Annethe Moe Jensen. 2020. The Role of Explicit Instruction on Students Learning Their Second Programming Language. In *United Kingdom and Ireland Computing Education Research Conference.* (Glasgow, United Kingdom) *(UKICER '20).* Association for Computing Machinery, New York, NY, USA, 10–16. https://doi.org/10.1145/3416465.3416475

[159] UKACMSIGCSE. 2021. *United Kingdom Computing Education Practice conference.* UKACMSIGCSE, New York, NY, USA. Retrieved November 1, 2021 from https://cepconference.webspace.durham.ac.uk

[160] Leo C Ureel, James Heliotis, Mohsen Dorodchi, Mireilla Bikanga Ada, Victoria Eisele, Megan E Lutz, and Ethel Tshukudu. 2020. Directing Incoming CS Students to an Appropriate Introductory Computer Science Course. In *2020 IEEE Frontiers in Education Conference (FIE).* IEEE, IEEE, New York, NY, USA, 1–9. https://doi.org/10.1109/FIE44824.2020.9274037

[161] Ian Utting, Neil Brown, Michael Kölling, Davin McCall, and Philip Stevens. 2012. Web-scale data gathering with BlueJ. In *Proceedings of the ninth annual international conference on International computing education research.* ACM, New York, NY, USA, 1–4.

[162] David W Valentine. 2004. CS educational research: a meta-analysis of SIGCSE technical symposium proceedings. *ACM SIGCSE Bulletin* 36, 1 (2004), 255–259. https://dl.acm.org/doi/pdf/10.1145/1028174.971391

[163] Lev Semenovich Vygotsky. 1980. *Mind in society: The development of higher psychological processes.* Harvard university press, Harvard, MA, USA.

[164] Susan Wiedenbeck. 1988. Learning recursion as a concept and as a programming technique. *ACM SIGCSE Bulletin* 20, 1 (1988), 275–278.

[165] Niklaus Wirth. 1984. From programming language design to computer construction. In *ACM Turing award lectures.* Association for Computing Machinery, New York, NY, USA.

[166] Niklaus Wirth. 2002. Computing science education: the road not taken. *ACM SIGCSE Bulletin* 34, 3 (2002), 1–3.

[167] Michael J. Wise. 1996. YAP3: Improved Detection of Similarities in Computer Program and Other Texts. In *Proceedings of the Twenty-Seventh SIGCSE Technical Symposium on Computer Science Education* (Philadelphia, Pennsylvania, USA) *(SIGCSE '96).* Association for Computing Machinery, New York, NY, USA, 130–134. https://doi.org/10.1145/236452.236525

[168] Peter Woit. 2011. *Not even wrong: The failure of string theory and the continuing challenge to unify the laws of physics.* Random House, London, England, UK.

[169] David Wood, Jerome S Bruner, and Gail Ross. 1976. The role of tutoring in problem solving. *Journal of child psychology and psychiatry* 17, 2 (1976), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

[170] David Wood and David Middleton. 1975. A study of assisted problem-solving. *British Journal of Psychology* 66, 2 (1975), 181–191.

[171] Cheng-Chih Wu, Nell B Dale, and Lowell J Bethel. 1998. Conceptual models and cognitive learning styles in teaching recursion. In *Proceedings of the 29th SIGCSE technical symposium on Computer science education.* Association for Computing Machinery, New York, NY, USA, 292–296.

[172] Daniel Zingaro and Leo Porter. 2015. Tracking Student Learning from Class to Exam Using Isomorphic Questions. In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education* (Kansas City, Missouri, USA) *(SIGCSE '15).* Association for Computing Machinery, New York, NY, USA, 356–361. https://doi.org/10.1145/2676723.2677239