CHAPTER SEVEN

MAKING AN IMPACT: FIT-FOR-PURPOSE CREATIVITY ASSESSMENT

DAVID H. CROPLEY

Abstract

The field of creativity research, in its modern sense, has existed for some 70 years. Initially driven by questions anchored in education, creativity research has branched out over the last 70 years to touch on many areas of human activity. However, it can be argued that the discipline has, in those 70 years, failed to make a deep and lasting impact in areas such as education. Despite strong interest in creativity, many countries still struggle with the issue of how to develop and assess creativity across the range of disciplines.

A key cause of this gap between research and practical application, especially in education, but also in business, may be a problem of measurement. In particular, the issue is a matter of the *fitness-for-purpose* of creativity measurement. This can be understood in terms of five key factors: (a) domain-specificity; (b) consistency and trustworthiness; (c) classroom integration; (d) speed of results, and (e) cost.

The good news is that solutions are now emerging to these problems in relation to creativity assessment. Computational methods, especially, but not limited to, the broad field of Artificial Intelligence (AI) are now becoming available. Already, these methods are demonstrating that they can solve many of the weaknesses identified here.

This chapter will delve into the application of computational methods to creativity measurement, giving examples of existing work in this area, and explaining why this is so important as a means of addressing the gap between creativity research and its real-world application. In an era where creativity is increasingly accepted as a vital 21st century competency, these computational methods could not have arrived at a better time.

Key Ideas/debates and categories addressed:

Debate: What innovations are currently in process?

Categories: Business/Education/Research

Keywords: Creativity, measurement, education, application, fitness-for-purpose, computational methods, artificial intelligence.

1. Creativity Research and Impact: Phase 1 (1950-2011)

The purpose of research is to increase understanding of phenomena of interest and originates with a question or problem (e.g., Leedy & Ormrod, 2013, p. 2). The practical, problem-focused nature of research leads many to regard all research as *applied*. Indeed, Brown et al (1997) quote J F Lovering[1] who said that "research is of two types: applied and yet to be applied" (p. viii). Inevitably, this means that a key metric for applied research is *impact*: what does the research contribute to the wider society and economy? Creativity research is no different: what contribution has creativity research made to societies more broadly since the beginning of the modern creativity era?

1.1 A Solution Waiting for a Problem?

Creativity, in its modern, psychological sense, has been a subject of scientific inquiry for over 70 years. Creativity research in *all* contexts has grown strongly over the same period. Since 1950[2], the volume of creativity research[3], as evidenced by the number of articles published (Figure 1), has risen from 460 articles (1950) to 148,000 articles (2020). This average growth rate of nearly 8%, year on year (not withstanding some recent anomalies, at least partly attributable to COVID), would suggest a sustained and healthy interest in the broad topic.

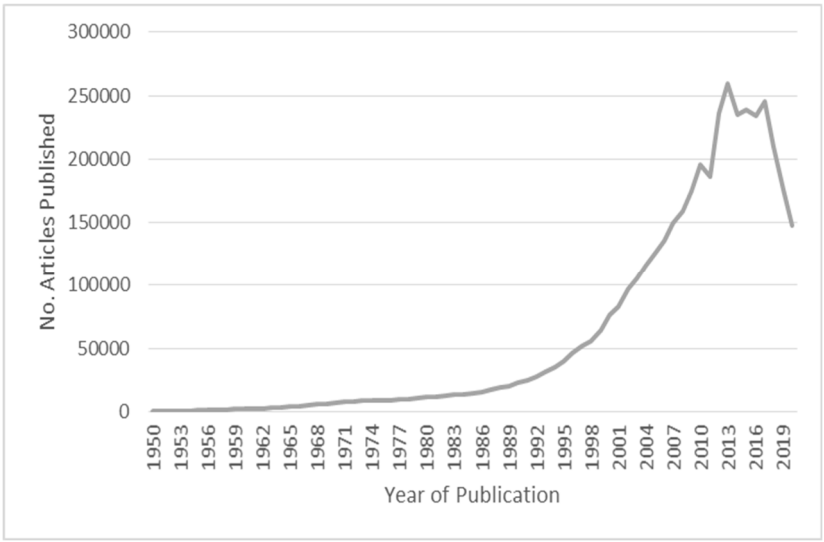

Figure 1: The Growth of Published Articles on *Creativity* (1950-2020)

[1] J F Lovering is a former Vice Chancellor (President) of Flinders University in Adelaide, South Australia.
[2] 1950 is often regarded as the beginning of the *modern era* of creativity research (e.g., Cropley & Cropley, 2013, p. 10), stimulated by Guilford's (1950) famous article.
[3] The metric I have used here is the number of articles reported, by Google Scholar, with the word "creativity" *anywhere in the article*, for the single years in question. The data presented here were collated in October 2022.

More specifically, much of the focus of creativity research has been directed towards questions of education. Indeed, an important, early driver of the modern creativity era was the role of creativity as a component of intelligence, and its role in the context of school education (e.g., Guilford, 1950). The intersection of creativity and education has remained a major focus of creativity research ever since, comprising the majority of published articles in creativity for almost all of the modern creativity era (Figure 2).

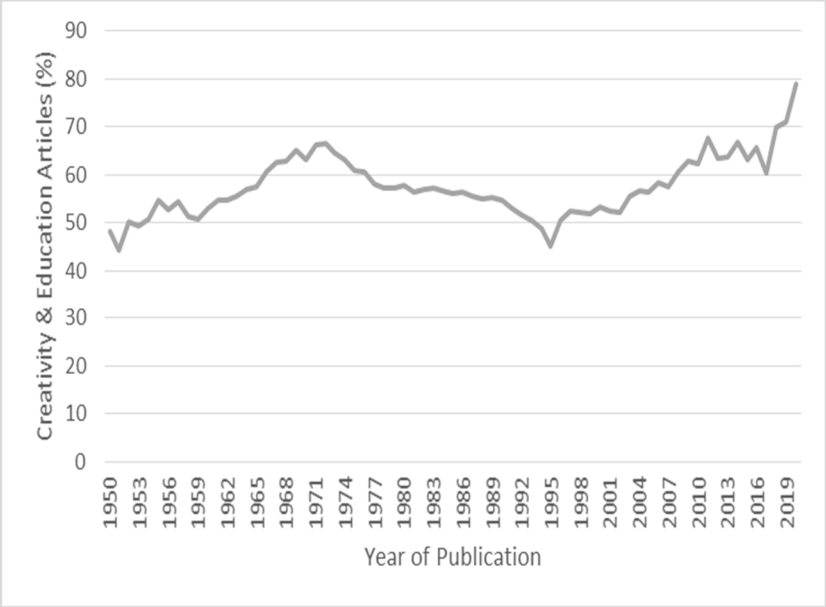

Figure 2: "Creativity & Education" as a Proportion of All Published Creativity Articles (1950 – 2020)

If we were to regard creativity *research* as a significant driver, or cause, of its application (in other words, more creativity research leads to more application of creativity research), then we might reasonably expect to see tangible evidence of the contribution, or *impact*, of this creativity research, especially in education (perhaps with a slight time lag). We can speculate that one way the impact of creativity in education should be seen is in the school and university curricula of various countries. Put simply, research demonstrates, with the support of empirical evidence, the importance of creativity in education more generally (e.g., Fasko, 2001) and in specific ways (e.g., Kaufman et al, 2021). Stakeholders in education (i.e., end-users such as school administrators, teachers, accreditation authorities, professors) notice this and react accordingly, building creativity into the curricula of their respective education systems. The evidence of impact therefore should be clear and simple. Creativity should, after 70 years and thousands of research articles, be firmly embedded in school and university curricula around the world. That fact that this uptake, even in 2021, is neither consistent, nor comprehen-

sive (see, for example, the discussion on creativity in school curricula in Patston et al, 2021, or the broader discussion, including higher education, in Cropley & Cropley, 2009) is, prima facie, evidence of a significant *disconnect* between creativity research and end-user application (or *impact*). Indeed, surveys of employers consistently back up this disconnect, complaining of skill deficiencies in creativity among university graduates (see Cropley, 2015 for a longer discussion). Why has creativity research – much directed at questions of education – *not* had a greater end-user impact in education? Why has it not made a more tangible contribution?

One possible explanation for the disconnect is that creativity research, in terms of application and impact, largely has operated (consciously or unconsciously) on a model of *technology push*, or what could be described as the *Field of Dreams* approach to research application and impact (i.e., "If you build it, they will come."). This approach assumes that the potential end-users of creativity research – e.g., parents, teachers, school or college administrators, businesses leaders – will recognise the inherent value of the body of knowledge that researchers are building and will adopt it as a solution to a problem that they may not even know they had. There are, of course, two risks to this supply-driven approach to research. The first risk is that the potential end-users of the research fail to perceive a problem for which creativity research is a solution. The second risk is that other problems, perceived as more urgent and compelling by potential end-users, occupy their finite resources and attention. In terms of end-user engagement and impact, it could be said, therefore, that creativity researchers, over many years, developed a solution that, at least from the evidence, was (or perhaps still is) *waiting for a problem*. What other evidence, from the end-user side of the equation, might support this *technology push* hypothesis of creativity research application and impact?

Early in the modern creativity era, Getzels and Jackson (1962) reported evidence that teachers struggled to identify creative students. Here, perhaps, was an early catalyst for research impact, and an opportunity for creativity researchers to respond to an end-user need. However, only modest, linear growth in creativity research over the following decade addressing this *measurement* and *education* focus[4], coupled with the comments of Wakefield (1987) that "there are a fair number of creativity tests on the market, [but] only a few have significant psychometric support, and most are recommended *for research purposes only*" (p. 19), suggests that little real progress was made in responding to the needs of end-users in education. Indeed, in the decade following Getzels and Jackson's (1962) book, the proportion of creativity research focused on *measurement* in education fell relative to the total volume focused on education (Figure 3) on the next page.

[4] Number of published articles containing the keywords "creativity" and "education" and "measurement".

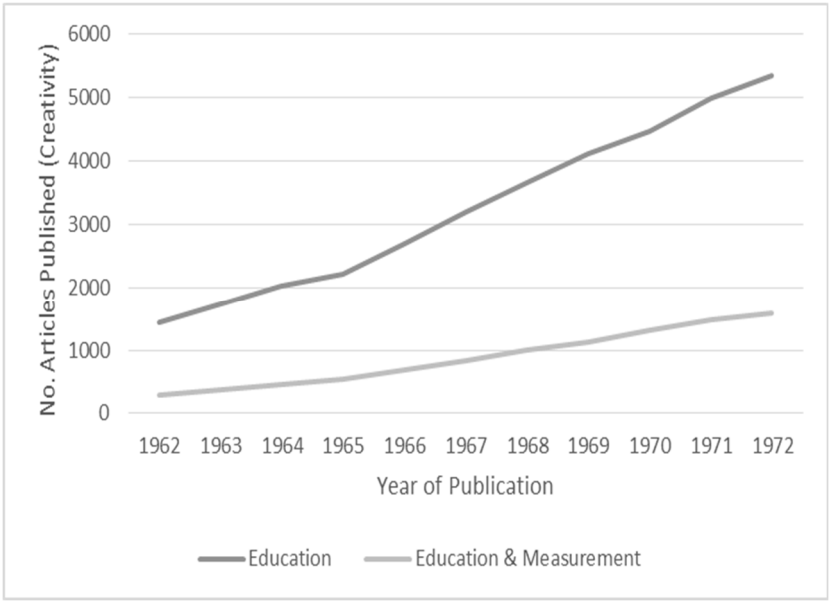

Figure 3: The Growth of a Focus on *Measurement* in Creativity & Education (1962 – 1972)

Even when a specific, compelling problem was identified, the opportunity, decisively, to bridge the disconnect – to make a decisive contribution – was not taken, with creativity research struggling to provide a solution. Two publications addressing the relationship between college admissions testing and creativity illustrate this. Sternberg (2010), for example, noted a variety of issues and weaknesses in relation to college admissions testing (keeping in mind that this discussion is very US-centric). He identified the nature of the problem: US college admissions testing does not test explicitly for creativity, and in failing to do so, may well be biased against the creativity of potential students. That there may be many reasons for this, not least the fact that college admissions testing in the US is big business, and that testing companies may be reluctant to do anything that disturbs their lucrative monopoly, is noted. However, what stands out is also the failure to proffer a concrete solution. Sternberg reiterated that a problem exists, but no real solution was offered, except to say that creativity *could be* a part of college admissions testing.

Dollinger (2011), in similar fashion, explored the relationship between creativity and college admissions testing. While both sides of the underlying argument are discussed – do current admissions tests address creativity or do they not – we end up with a disconnect. The same end-user problem (how to measure creativity explicitly in college admissions processes) was identified, calls were made to address this gap ("…alternative assessments should be used if admissions committees wish to select those with the greatest creative potential.", p. 337), but nothing tangible was really offered as a solution to the problem. Indeed, the measures of creativity used by Dollinger

(2011) to show that the ACT[5] has a modest ability to predict creativity are no practical solution to this problem, being either a self-report measure of "creative" accomplishments based on Hocevar (1979), a non-standard version of the TCT-DP (Urban & Jellen, 1996) scored by up to eight quasi-expert judges, or a photographic homework assignment scored by up to five non-expert judges.

To be fair, these calls for change may have had some small effect in closing gap between creativity research and end-user impact, at least in education. Creativity research focused on measurement and education, in the decade from 2010, grew as a proportion of all research in creativity and education (Figure 4). Nevertheless, the growth of creativity research in education, focused on measurement, has been modest for such a compelling end-user problem. Equally, simply doing more applied research does not guarantee that the research gets applied.

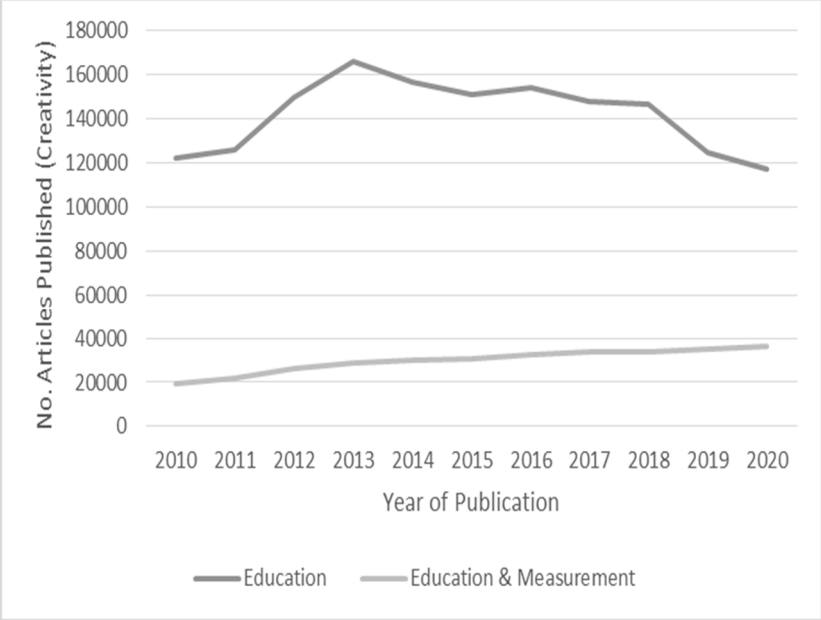

Figure 4: The Growth of Creativity Research (>2010) focused on Education and Measurement

This state of affairs, namely, the apparent reluctance of creativity researchers to address end-user needs, is reminiscent of a broader phenomenon known as Cobb's Paradox[6]. Stated in terms of creativity research over the first phase of

[5] The paper, unfortunately, never actually says what the "ACT" is. The implication is that it is some sort of test used for college admissions in the US, without ever giving any specifics.

[6] Martin Cobb, Chief Information Officer (CIO) for the Secretariat of the Treasury Board of Canada stated, in 1995, "We know why projects fail, we know how to prevent their failure -- so why do they still fail?"

the modern era (1950-2011), we see a variant of Cobb's Paradox: *we know that creativity is important to education, we know how to measure creativity – so why are measures of creativity used so little in education?*

Perhaps the long-standing disconnect between creativity research and end-user impact was, and still is, driven by Wakefield's (1987) observation that "…no current measure of creativity is adequate." (p. 18)? Perhaps the driver of the gap between creativity research and end-user application has been a long-standing problem of measurement? In fact, Kaufman (2010), also discussing the issue of creativity in college admissions testing, gets to the real source of the disconnect. Not only does he note (discussing the publishers of standardised tests) that "If it were easy to add a construct like creativity to a standardised test, it would be done" (p. 197), but he also states that "Creativity simply does not yet have an affordable, valid, and easy-to-administer large-scale measure that is not susceptible to coaching and faking" (p. 197). Indeed, Kaufman (2010) goes on to point out the specifics of this measurement problem: questionable validity, inconsistency, complexity, and cost. In simple terms, the heart of the disconnect between creativity research and end-user application may be the fact that *good* measures of creativity are simply not *fit-for-purpose* from the point of view of end users.

Most worrying of all, a consequence of the disconnect between creativity research and end-users – the lack of fit-for-purpose measures of creativity – may be the cause of a decline in research in educational aspects of creativity (see Figure 4) just at a time when the *need* for creativity in education is growing.

2. Creativity Research and Impact: Phase 2 (2011 – 2017)

Whatever the historical relationship between creativity research and impact, there is a sense that something has begun to change in the broader environment. That change, in turn, may be creating conditions in which it will be easier for creativity research to make an impact. What is the change, and how is that affecting creativity research and its potential for end-user impact?

Digitalisation and the Future of Work

Beginning in around 2011, a decisive shift in the relationship between creativity research and end-user application began to take place. The rapid *digitalisation* of societies, dubbed Industry 4.0 by the German government in that year (see Cropley & Cropley, 2021), began to influence, more and more, how governments and organisations perceived the nature and value of the skills of human workers. There began, in other words, a recognition that in a world increasingly characterised by big data, artificial intelligence and automation, the skills that humans would need were changing (e.g., OECD, 2017). The jobs of the future, so the argument goes, would increasingly focus on skills that were *uniquely* human – skills for which humans could not be replaced by machines – and creativity was recognised as such a skill. The effect of this change in thinking was to alter the relationship between creativity researcher and end-user, from technology push to *market pull*. In a very short space of time, the balance in the relationship between creativity researchers and end-

users shifted to a situation in which end-users realised they had a problem – how to build creativity into education – and began actively seeking solutions.

The growing end-user demand for creativity, since the advent of digitalisation, is readily apparent. In 2010, the Australian Curriculum, Assessment and Reporting Authority[7] (ACARA), for example, first introduced critical and creative thinking into the national curriculum, as a *general capability*, from kindergarten to grade 10. Beginning in 2013 the World Economic Forum[8] began a process that would result, in early 2016, in the first *Future of Jobs* report (WEF, 2016). That report, updated annually since 2016, has consistently highlighted creativity, among a variety of factors, as a core *21st century skill*. The Organisation for Economic Cooperation and Development (OECD), through its Program for International Student Assessment (PISA), in 2022 introduced, for the first time, a test of creative thinking as part of its suite of assessments[9].

Thus, it would appear, the conditions for bridging the disconnect between creativity research and end-user application were, beginning around 2011, at hand. Digitalisation had created a compelling need for creativity in education, and creativity research was ideally placed to support this need. Indeed, from 2011 to 2013, there was a marked increase in creativity research outputs, both in creativity generally, and in creativity and education more specifically (Figure 5).

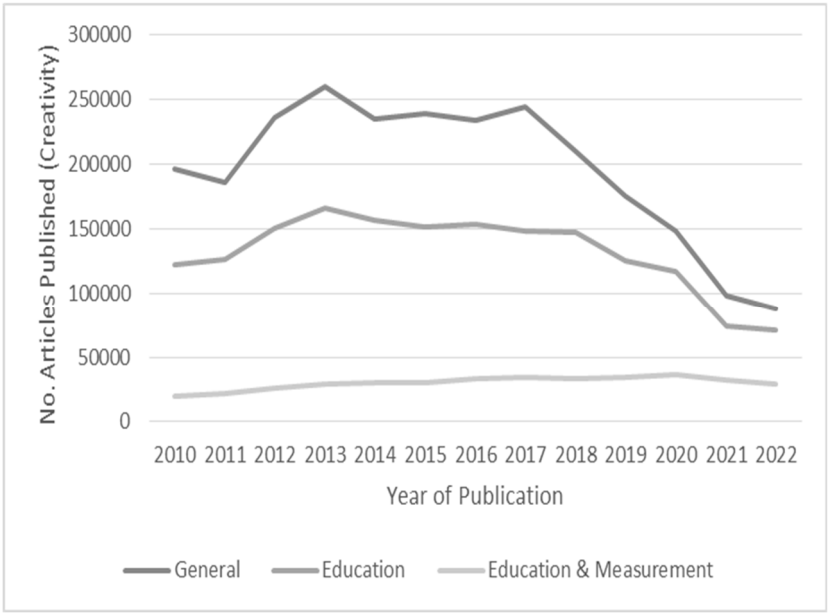

Figure 5: Changes to Creativity Research Output Post 2011

7 ACARA is the Australian Governments national
8 See: https://www.weforum.org/reports/the-future-of-jobs-report-2020/
9 See: https://www.oecd.org/pisa/innovation/creative-thinking/

It is therefore altogether more surprising that, in 2013, research outputs in creativity (both generally and in education: Figure 5) began to fall from an all-time peak. In the presence of strong, indeed growing, end-user demand, could this decline, particularly sharp since 2017, be explained by the failure of creativity research to respond to the demand? Was the lack of fit-for-purpose measures of creativity not only failing to meet end-user demand, but turning researchers, frustrated by their inability to meet the demand, away from creativity research? Or (Figure 5) was the end-user demand causing a small, but significant, reorientation of creativity research *into* questions of education and measurement, especially since 2017?

3. Creativity Research and Impact: Phase 3 (2017+)

Digitalisation has refreshed the impetus for impactful creativity research. It has created a compelling demand. Although evidence seems to suggest that overall creativity research output is falling, despite this increased demand, the reality may be that creativity research is actually beginning to focus better on end-users and impact. However, even as creativity research undergoes a more impact-focused reorientation, there are other external forces at work that complicate research impact.

Reorientation?

Demand for creativity has never been stronger, thanks to digitalisation and a strong end-user need. However, since early in this era of digitalisation, creativity research in general, and in education, has been shrinking (Figure 5). I have argued that this may be due, at least in part, to creativity research's *Achilles' heel*: a broad lack of fitness-for-purpose in creativity measures. However, this weakness in the translation of research to end-user impact may not be unique to *creativity* research.

Much attention has been given, recently, to the so-called *replication crisis* in psychological research (e.g., Chambers, 2019). Usually explained in terms of structural pressures (e.g., the pressure on academics to *publish or perish*; the *publication bias* that favours statistically significant results) and the Questionable Research Practices (QRPs) that may result (e.g., exploiting researcher degrees of freedom, selective reporting, HARKing) it seems that surprisingly little attention has been given to the *quality of measurement* of many psychological constructs as a possible cause.

It is easy to see why this might be the case. The replication crisis is an embarrassing failure for the field of psychology. Far better, then, to focus on *external* causes. The poor level of replication results from poor practices that are forced on innocent psychologists against their will, with the misbehaviour of just a few bad apples also adding to the problems. On the other hand, to attribute the replication crisis to poor psychometric measurement strikes much more directly at the heart of the discipline. It would be necessary to admit to a systemic, *internal* failure – much closer to incompetence – than to admit to a reaction to an unfair system, or the misconduct of a minority.

In fact, the notion that the *quality of measurement* in psychological research might be, at least partly, at fault has not been entirely overlooked. Fried and Flake (2018) explore this line of reasoning, noting that the cultural shift underpinning open science *has largely ignored the topic of measurement* (p. 1), and stressing that *the quality of measurement is even more foundational than statistical practice* (p. 1).

Is the underlying cause of the replication crisis more generally, and the lack of impact in creativity research more specifically, fundamentally a matter of *poor-quality*[10] measurement? Is poor-quality measurement itself a reflection of a system of research driven not by end-user need (i.e., *market pull*), but by research, in effect, for the sake of research (i.e., *technology push*)?

Whatever the underlying causes, one thing is clear. There is now, in creativity research, an unequivocal end-user demand in education, driven by digital transformation and the future of work. Creativity research has an opportunity – perhaps only fleeting – to respond to this need, but to do so we must first understand the symptoms of poor-quality measurement – especially fitness-for-purpose – before addressing how these symptoms might be alleviated. To succeed offers the prospect of a new era of highly impactful creativity research, central to preparing students to thrive in the era of the Future of Work. To fail risks the future prosperity of current students who will be entering a workplace where creativity is at a premium. To fail also risks permanent damage to the reputation of the discipline of creativity research.

A Glimmer of Hope

In fact, this process – addressing fit-for-purpose measurement in creativity research – may be already underway. The volume of creativity research specifically in education and measurement has, in defiance of the broader trend, grown since 2010 (and ignoring the likely impact of COVID in 2021 and 2022) as suggested by Figure 6. Furthermore, in 2022, the proportion of publications in creativity and education has grown to its highest ever level (80% of all creativity outputs: Figure 2).

[10] I use this term very broadly to encompass poor fitness-for-purpose as well as other factors such as validity and reliability.

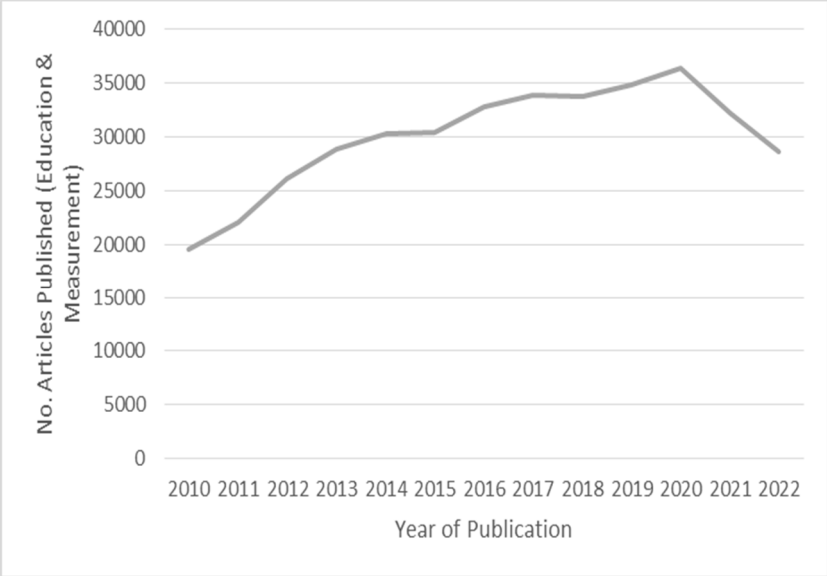

Figure 6: The Growth of Research (>2010) focused on Measurement and Education in Creativity

The following sections therefore explore the symptoms of poor fitness -for-purpose of measurement in creativity, before examining how creativity research may be tackling this issue. Although it is perhaps too early to draw firm conclusions as to the cause of this positive trend, it may well be that a solution to the measurement problem in creativity research is already making itself felt.

4. Fitness-for-Purpose and Measures of Creativity

A growing focus on the fitness-for-purpose of creativity assessments highlights additional weaknesses driven by end-users' (i.e., teachers, students) needs in educational settings. Race (2014) outlined criteria that define good (i.e., *fit-for-purpose*) measures in education that can be applied to measures of creativity. To be fit-for-purpose, measures should be: (a) valid (they should measure what they are intended to measure); (b) reliable (they should be consistent and fair); (c) authentic (they should be connected to real-life application and be meaningful to the student), and (d) transparent (they should be matched to subject descriptions and outcomes). Additionally, fit-for-purpose assessments should facilitate both formative and summative assessment (Burke, 2010) balancing the need for both criterion- and norm-based assessment contexts (Biggs & Tang, 2011). Driven by the forces of digitalisation and the future of work, and given the focus on creativity as a 21st-century competency, these criteria help to answer the question: Are current creativity assessments fit-for-purpose in educational settings?

To answer this question, we first set aside the broad technical and psychometric issues raised by creativity scholars (e.g., Plucker, Makel, & Qian, 2019), and discussed in Barbot, Hass, & Reiter-Palmon, 2019). From the point of view of *end-users* in education, and consistent with the theme of impact, current creativity assessments lack fitness-for-purpose because:

1. They fail to allow for creativity assessment tailored to specific domains (e.g., science vs art).
    a. If creative performance, as many researchers suggest, has limited domain-generality (see, for example, Barbot, Hass, & Reiter-Palmon, 2019, p. 234; Sternberg, 2020, p. 21), then creativity assessments in educational settings must reflect this. Creativity assessments specific to a diverse range of individual subjects must be made available to end-users.
2. They are inconsistent and therefore untrustworthy.
    a. Even creativity tests with a strong reputation for objectivity, reliability and validity (e.g., the Test of Creative Thinking – Drawing Production, created by Urban & Jellen, 1996) are susceptible to the variability associated with human raters. The TCT-DP, for example, demonstrates typical inter-rater reliability values of .90-.95 (e.g., Theurer, Berner, & Lipowsky, 2016). Even though this is considered *high* in a statistical sense, it reflects a level of inconsistency that impacts the fitness-for-purpose of creativity assessments. The more *subjective* creativity assessment methods (for example, the *Consensual Assessment Technique*, Amabile, 1982) are even more prone to this erosion of trust in creativity assessment in end-user contexts. Dollinger, Urban and James (2004), for example, noted an inter-rater agreement of just .78 using the CAT. The impact of this form of subjectivity is discussed by Beaty and Johnson (2021) in more detail. For creativity assessments to be widely adopted in schools they must be trusted, especially in high-stakes settings such as accountability systems or college admissions (Plucker, Makel, & Qian, 2019, p. 59). To be trusted, inconsistencies among human judges must be eliminated. Inter-rater *dis*agreements are akin to *misdiagnoses* and may lead to faulty, high-stakes decisions.
3. They fail to integrate, seamlessly or authentically, into the classroom context.
    a. General tests of creative potential (for example, *divergent thinking* operationalized in the Alternate Uses Test, Torrance, 1988) do not integrate smoothly with activities in most classroom subjects. In a mathematics class, for example, how would a test of ideational originality (e.g., *how many different uses can you think of for a brick?*) relate to a student's potential or performance in mathematical creativity, especially given the known difficulty that such tests have in correlating originality across two different stimuli

(e.g., Reiter-Palmon, Forthmann, & Barbot, 2019)? Many (perhaps most) current creativity assessments lack *authenticity* (Race, 2014) and force teachers to interrupt the flow of student learning in order to administer these inauthentic tests.

4. They do not provide rapid results/feedback.
    a. Many creativity tests are relatively quick and easy to administer regardless of whether they are assessments of creative potential or performance, or self-report measures. However, the critical issue for end-users is the speed with which results are available. The longer the gap between conducting the test and obtaining the results of creativity assessments, the lower the utility of those results. Formative feedback, in particular, requires fit-for-purpose creativity assessments that give end-users the flexibility to incorporate these assessments, with little or no delay, into the learning process (Burke, 2010).
5. They are expensive to use.
    a. Whether objectively defined test of creative ideation, assessment of related personality traits (e.g., Openness to Experience), or subjective measure of creative performance, creativity assessments usually involve what Beaty and Johnson (2021) define as *effort*. This effort – in other words, the human labour required to administer, score, analyse and report the assessments – results in creativity assessments that are expensive to use. This cost creates a barrier that leads to one of two outcomes. In the best case, end-users adopt methodologically weak measures because they are cheaper (e.g., Reiter-Palmon, Forthmann, & Barbot, 2019). In the worst case, end-users abandon the use of creativity assessment altogether. The paradox of creativity assessments is that good quality comes at a cost, and high cost deters end-users.

Although these criteria present an unfavourable analysis of current creativity assessment, the symptoms described serve one very important purpose. Together, they comprise a set of design criteria for novel, fit-for-purpose approaches to creativity assessment. Indeed, these criteria echo characteristics that have been called for previously in impact-focused scholarship (e.g., Lucas, Claxton, & Spencer, 2013). However, as discussed earlier, there is a glimmer of hope for creativity assessment. The recent, growing proportion of creativity research focused on measurement in the context of education (Figure 6) contains promising examples of new approaches to *fit-for-purpose* creativity assessment, focused on end-users. Broadly speaking, these are the development of *computational* approaches to creativity assessment that have begun to emerge only very recently. The final section of this chapter will outline some of the key computational methods and techniques that are shaping the development of fit-for-purpose measures of creativity.

5. Computational Methods Applied to Creativity Measurement

Computational approaches to creativity, in the broadest sense, are not new. From very early in the modern creativity era, researchers asked how creativity might be modelled, or even replicated, by computational systems (e.g., Newell, Shaw, & Simon, 1962). Bruner (1962), however, believed that *artificial* creativity would need to consist of more than just the automation of the *blind generation* of alternatives. Notwithstanding the dormant periods (or *winters*) of AI research, arguably driven by the inability of technology, tools and methods to deliver promised advances, Boden (1998) continued the scholarship of creativity and artificial intelligence (AI). While it is true that some efforts were made, as far back as the 1980s, to automate aspects of creativity (or related) assessment (e.g., Simonton, 1986, 1990) these approaches used computers only, for example, to *assist* in activities such as content analysis of blocks of text. Over the decades since the 1960s, research efforts in computational creativity seem to have focused onto three elements: (a) building an artificial system that is capable of creativity; (b) defining algorithmic descriptions of human creativity, and (c) developing artificial aids to human creativity (see Cropley, Medeiros, & Damadzic, 2022). Within this framework, many scholars (e.g., Wiggins, 2006; Colton & Wiggins, 2012) focus, in particular, on the idea of replicating human creative *behavior*.

It is notable, however, that across most of the history of computational creativity there is almost a complete absence of creativity *assessment* from the discussion. Boden (1998) attempted to explain why, noting (p. 347) that "…AI will have less difficulty in modelling the generation of new ideas *than in automating their evaluation*" [emphasis added]. Little, if any, research has attempted to apply computational approaches such as AI to the assessment of any aspect of creativity, in any meaningful way, until recently.

There is a certain irony that the same fundamental driver of digitalization and the future of work – Artificial Intelligence – that is causing the reorientation in creativity research (see the earlier discussion) is also the solution to the creativity measurement, *fitness-for-purpose* problem. What is rapidly becoming the dominant such approach centers on the concept of natural language processing (NLP) and, especially, Latent Semantic Analysis (LSA). Notable examples and discussions include Forster & Dunbar (2009), Dumas and Dunbar (2014), Harbinson and Haarmann (2014), Beaty and Johnson (2021), Altindis (2022), Acar (2021), Forthmann and Doebler (2022), Beaty et al (2022), Weinstein et al (2022) and Plucker (2022). Each of these used LSA as the basis of their approach to automate divergent thinking-based tasks such as the Alternate Uses Task (AUT).

Olson et al. (2021), however, is significant, not because the underpinning technique is different, but because those authors have attempted to bridge the vital gap between *applied research*, and *research that gets applied*, that is critical to the fitness-for-purpose of creativity assessments. While their verbal *Divergent Association Task* (DAT) does not necessarily address issues such as domain diversity and integration into the classroom (Race, 2014), it makes significant inroads into consistency and trust, speed and cost.

Other computational (AI) approaches to creativity assessment are also emerging rapidly. These have the potential to address not only trust,

speed and cost, but also the need for a diversity of domain coverage as well as seamless classroom integration. For example, Kovalkov et al. (2022) uses a machine learning approach to assess the creativity of computer programs, while Marrone, Wang, and Cropley (2022) apply natural language processing and concepts of latent semantic analysis to the assessment of mathematical creativity. Also important, particularly in terms of breaking away from an over-reliance on LSA and therefore a purely verbal operationalization of divergent thinking, is the recent work of Cropley and Marrone (2022), who used a Computational Neural Network (CNN) to automate the assessment of the (figural) Test of Creative Thinking – Drawing Production (Urban & Jellen, 1996) with high accuracy.

The advances that have been made recently are important and go a long way to addressing long-standing issues of fit-for-purpose creativity measurement specifically, and impact more generally. However, weaknesses remain in many of these novel approaches. While Beaty and Johnson (2021) and Olson et al. (2021) address issues in consistency (subjectivity) and effort (speed and cost), they may still retain weaknesses concerning domain-specificity and authenticity. If Cropley and Marrone (2022) and Kovalkov et al. (2022) are better focused on the assessment needs of end-users, they too retain weaknesses regarding domain-specificity or authenticity. What is clear is that many important advances are being made – all-based around AI and machine learning (ML) – that have the potential to provide end-users with truly fit-for-purpose creativity assessments.

6. Conclusions

Creativity research has struggled with impact for much of the modern era. However, since 2011, a new external driver – digitalisation and the impact of this on the future of work – has been the catalyst for a reorientation in creativity research. End-users are now demanding fit-for-purpose measures of creativity that will help deliver a workforce equipped with the competencies needed to thrive in a world where AI has taken over many of the routine, algorithmic physical and cognitive tasks that humans previously performed. Key to this reorientation has been the ability of creativity research to respond to the need for fit-for-purpose measures of creativity.

The effective response is not the development of *new* measures of creativity, but the *automation* of existing measures of creativity, typically using the tools and methods of AI including machine/deep learning. However, even as this new era of creativity assessment gathers pace, and the problems of end-user impact begin to lbe tackled, there is a danger that the focus again turns away from end-users. The rapid growth of articles exploring latent semantic analysis (LSA) as a method for assessing verbal divergent thinking must not lose sight of the end-user. Neither divergent thinking, nor verbal divergent thinking, are the entirety of creativity. While these measures may address some of the elements of fitness-for-purpose (e.g., speed, cost), they do not address all of them. Indeed, no matter how well creativity researchers address some of the technical issues of measurement, if we are not *measuring the right things*, from the point of view of end-users, we may, in the end, return to another period of low impact.

If creativity research is to deliver high end-user impact, as all applied research should, then the field must redouble its efforts to apply computational approaches to the broadest possible range of highly reliable and valid creativity measures. Automation of existing measures may be the most cost-effective and practical approach. The tools to achieve this are now available.

References

Acar, S., Berthiaume, K., Grajzel, K., Dumas, D., Flemister, C. T., & Organisciak, P. (2021). Applying Automated Originality Scoring to the Verbal Form of Torrance Tests of Creative Thinking. *Gifted Child Quarterly*. doi:https://doi.org/10.1177/00169862211061874

Altindiş, Z. T. (2022). Perspective Chapter: New Approaches to the Assessment of Domain-Specific Creativity. In *Creativity*: IntechOpen.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of personality and social psychology, 43*, 997-1013.

Barbot, B., Hass, R. W., & Reiter-Palmon, R. (2019). Creativity assessment in psychological research:(Re) setting the standards. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 233-239.

Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior research methods, 53*(2), 757-780.

Beaty, R. E., Johnson, D. R., Zeitlen, D. C., & Forthmann, B. (2022). Semantic Distance And the Alternate Uses Task: Recommendations for Reliable Automated Assessment of Originality. *Creativity Research Journal, 34*(3), 245-260. doi:https://doi.org/10.1080/10400419.2022.2025720

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university*. Maidenhead, UK: McGraw-Hill International.

Boden, M. A. (1998). Creativity and artificial intelligence. *Artificial intelligence, 103*(1-2), 347-356.

Brown, R. H., Carter, A. R., Davie, R. S., & Ogilvie, G. J. (1997). *Research, development and innovation.* Crows Nest, NSW: Engineers Australia.

Bruner, J. S. (1962). The Conditions of Creativity. In H. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary Approaches to Cognition* (pp. 1-30). New York, NY: Atherton Press.

Burke, K. (2010). *Balanced assessment: From formative to summative*. Bloomington, IN: Solution Tree Press.

Chambers, C. (2019). *The Seven Deadly Sins of Psychology*. Princeton, NJ: Princeton University Press.

Colton, S., & Wiggins, G. A. (2012). *Computational creativity: The final frontier?* Paper presented at the Ecai.

Cropley, A. J., & Cropley, D. H. (2009). *Fostering creativity: A diagnostic approach for education and organizations*. Cresskill, NJ: Hampton Press.

Cropley, D. H. (2015). Promoting creativity and innovation in engineering education. *Psychology of Aesthetics, Creativity, and the Arts, 9*(2), 161-171.

Cropley, D. H., & Cropley, A. J. (2013). *Creativity and crime: A psychological approach*. Cambridge, UK: Cambridge University Press.

Cropley, D. H., & Cropley, A. J. (2021). *Core Capabilities for Industry 4.0 - Foundation of the Cyber-Psychology of Engineering*. Bielefeld, Germany: Wbv Media.

Cropley, D. H., Madeiros, K., & Demadzic, A. (2022). Creativity and Artificial Intelligence: The Integration of Human and Artificial Cognition. In D. Henriksen & P. Mishra (Eds.), *Creative Provocations: Speculations on the Future of Creativity, Technology & Learning*. Cham, Switzerland: Springer.

Cropley, D. H., & Marrone, R. L. (2022). Automated scoring of figural creativity using a convolutional neural network. *Psychology of Aesthetics, Creativity, and the Arts*. doi:https://doi.org/10.1037/aca0000510

Dollinger, S. J. (2011). "Standardized minds" or individuality? Admissions tests and creativity revisited. *Psychology of Aesthetics, Creativity, and the Arts, 5*(4), 329-341.

Dollinger, S. J., Urban, K. K., & James, T. A. (2004). Creativity and openness: Further validation of two creative product measures. *Creativity Research Journal, 16*(1), 35-47.

Dumas, D., & Dunbar, K. N. (2014). Understanding fluency and originality: A latent variable perspective. *Thinking Skills and Creativity, 14*, 56-67.

Dunbar, K., & Forster, E. (2009). *Creativity evaluation through latent semantic analysis.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Fasko, D. (2001). Education and creativity. *Creativity Research Journal, 13*(3 -4), 317-327.

Forthmann, B., & Doebler, P. (2022). Fifty years later and still working: Rediscovering Paulus et al's (1970) automated scoring of divergent thinking tests. *Psychology of Aesthetics, Creativity, and the Arts, Online First*. doi:https://doi.org/10.1037/aca0000518

Fried, E. I., & Flake, J. K. (2018). Measurement matters. *APS Observer, 31*.

Getzels, J. A., & Jackson, P. W. (1962). *Creativity and intelligence: Explorations with gifted students*. New York, NY: Wiley.

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444-454.

Harbinson, J., & Haarman, H. (2014). *Automated scoring of originality using semantic representations.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Hocevar, D. (1979). *The Development of the Creative Behavior Inventory*. Paper presented at the Rocky Mountain Psychological Association.

Kaufman, J. C. (2010). Using creativity to reduce ethnic bias in college admissions. *Review of General Psychology, 14*(3), 189-203.

Kaufman, J. C., Kapoor, H., Patston, T., & Cropley, D. H. (2021). Explaining standardized educational test scores: The role of creativity above and beyond GPA and personality. *Psychology of Aesthetics, Creativity, and the Arts*.

Kovalkov, A., Paassen, B., Segal, A., Pinkwart, N., & Gal, K. (2022). Automatic Creativity Measurement in Scratch Programs Across Modalities. *IEEE Transactions on Learning Technologies*.

Leedy, P., & Ormrod, J. (2013). *Practical research: planning and design* (10 ed.). Upper Saddle River, NJ: Pearson Education Inc.

Lucas, B., Claxton, G., & Spencer, E. (2013). *Progression in Student Creativity in School: First Steps Towards New Forms of Formative Assessments*. Retrieved from http://dx.doi.org/10.1787/5k4dp59msdwk-en

Marrone, R., Cropley, D. H., & Wang, Z. (2022). Automatic Assessment of Mathematical Creativity using Natural Language Processing. *Creativity Research Journal*, 1-16.

Newell, A., Shaw, J. G., & Simon, H. A. (1962). The process of creative thinking. In H. E. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary Approaches to Creative Thinking* (pp. 63-119). New York, NY: Atherton Press.

OECD. (2017). *Future of work and skills*. Retrieved from https://www.oecd.org/els/emp/wcms_556984.pdf

Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., & Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences, 118*(25).

Patston, T., J., Kaufman, J. C., Cropley, A. J., & Marrone, R. L. (2021). What Is Creativity in Education? A Qualitative Study of International Curricula. *Journal of Advanced Academics*. doi:doi:10.1177/1932202X20978356

Plucker, J., Makel, M., & Qian, M. (2019). Assessment of Creativity. In J. Kaufman & R. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 44-68). Cambridge, UK: Cambridge University Press.

Plucker, J. A. (2022). The Patient is Thriving! Current Issues, Recent Advances, and Future Directions in Creativity Assessment. *Creativity Research Journal*, 1-13. doi:https://doi.org/10.1080/10400419.2022.2110415

Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 144.

Simonton, D. K. (1986). Aesthetic success in classical music: A computer analysis of 1935 compositions. *Empirical Studies of the Arts, 4*(1), 1-17.

Simonton, D. K. (1990). Lexical choices and aesthetic success: A computer content analysis of 154 Shakespeare sonnets. *Computers and the Humanities, 24*(4), 251-264.

Sternberg, R. J. (2010). *College admissions for the 21st century*: Harvard University Press.

Sternberg, R. J. (2020). What's wrong with creativity testing? *The Journal of Creative Behavior, 54*(1), 20-36.

Theurer, C., Berner, N. E., & Lipowsky, F. (2016). Assessing creative potential as student outcome: On the applicability of the TCT-DP in repeated measures. *Thinking Skills and Creativity, 20*, 74-82.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In R. J. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives* (pp. 43-75). New York, NY: Cambridge University Press.

Urban, K. K., & Jellen, H. G. (1996). *Test for Creative Thinking - Drawing Production (TCT-DP)*. Lisse, Netherlands: Swets and Zeitlinger.

Wakefield, J. F. (1987). *The Outlook for Creativity Tests*. Paper presented at the The Future of Special Education, Orlando, FL.

WEF. (2016). *The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution*. Retrieved from http:// reports.weforum.org/future-of-jobs-2016/

Weinstein, T. J., Ceh, S. M., Meinel, C., & Benedek, M. (2022). What's creative about sentences? A computational approach to assessing creativity in a sentence generation task. *Creativity Research Journal*. doi:DOI: 10.1080/10400419.2022.2124777

Wiggins, G. A. (2006). Searching for computational creativity. *New Generation Computing, 24*(3), 209-222.