Performance Anxiety in Academia: Tensions within research assessment exercises in an age of austerity

Dr. Charlotte Holland
School of Education Studies,
Dublin City University.

Dr. Tony Hall
School of Education,
National University of Ireland, Galway.

Dr. Francesca Lorenzi
School of Education Studies,
Dublin City University

Abstract.

The current recessionary economic climate in Ireland has (re-) awakened a neoliberal agenda that is changing the dynamic of what is being valued within research assessment exercises, specifically across Arts, Humanities and Social Sciences (AHSS) disciplines in higher education. Research assessment exercises in AHSS disciplines now place a greater emphasis on measuring performance in terms of quantitative research metrics (such as: *bibliometrics*, *impact factors* and/ or *citation indices*), in an attempt to demonstrate greater accountability and value-for-money within this age of austerity. This practice has the potential to impact negatively on the quality and diversity of research, as well as *'the autonomy of researchers*' (RIA, 2009, p2) undertaking AHSS research in Ireland and elsewhere. This article critically reviews research assessment exercises, with particular reference to the assessment of educational research in Ireland. It examines issues in the assessment of research within the neoliberal agenda that is evident in Ireland, and elsewhere. For example, in other jurisdictions, the neoliberal drive for accountability has been accompanied by an increase in 'citation clubs', a malpractice involving a group of researchers consistently citing each others work to increase their citation index. It also challenges the validity of utilising predominantly quantitative research metrics in light of the recent move towards online publication of research, where the manipulation of *meta-data* (key words that describe the research) has the potential to unfairly increase the citation indices of those researchers with a better understanding of *search optimisation techniques* within online contexts. The discussion concludes by critiquing some of the emerging and emergent anxieties in relation to researcher performance in assessment exercises.

Introduction

The current recessionary economic climate in Ireland has (re-) awakened a neoliberal agenda that is changing the dynamic of what is being valued within research assessment exercises, specifically across Arts, Humanities and Social Sciences (AHSS) disciplines in higher education. Economic recession has resulted in less public monies available for research, a more competitive environment within higher education, with an increased emphasis on justifying expenditure, improving quality and extending the impact of research. This is not a new phenonomen nor is it localised to Ireland. In 2007, the British Academy commented that management and researchers were coming under pressure from governments and funders to "*demonstrate that their results make a difference outside of a narrow academic context*" (p.5). In a speech delivered in Dublin City University on 11th September 2013, the Irish President Michael D. Higgins further highlighted the fascination with 'quantification' as a distinctive feature of our times, which he perceives as connected

to the ideological standpoint endorsed by neo-liberalists and strongly influencing governmental policies.

> *This question of measurement too is a vast and important one. I strongly believe that we need to re-examine the categories by which we gauge economic value and human worth, as well as the language we employ to do so. It is not an exaggeration to say that we live in times where economic worth is primarily seen as a matter of productive capacity.*

At a philosophical level, we are witnessing the confluence of two complementary tendencies in policies concerning research assessment in higher education: on the one hand the pervading neoliberal concern for productivity and on the other hand a modernist and scientific trust in quantification in the quest for transparent and accountable evidence of research outputs (Smith, 2010). The combination of these two tendencies has generated the requirement for greater transparency in research outputs, which in turn has resulted in assessing the quality of research work in terms of its quantitative evidence and of its verifiable socio-economic benefits.

While the demand for transparency and accountability is a justifiable, what is being counted as evidence of quality research is problematic on two grounds. Firstly, with Nussbaum (1990, p.36) - who refers to the tendency to reduce quality to quantity as '*ethical immaturity*'- it can be argued that not all valuable things are non-commensurable, thus a quantification of quality is doomed to failure or at least to approximation. Evidence generated through quantitative means can only be considered as partial and imperfect. Such imperfection becomes particularly apparent when considering the variety of research outputs produced in AHSS, which are difficult to evaluate using quantitative measures. Secondly, evidence of quality should be considered in relation to parameters of a normative nature, such as "originality, significance and rigour" as articulated within the Research Assessment Exercise (RAE) process in the United Kingdom. However, the preferred metrics of governments and higher education funding agencies are quantitative in nature, resulting in a shift in assessing the normative dimensions of research quality to assessing extrinsic features such as (i) the location of publications; (ii) the number of times they have been cited; (iii) the number of times they have been downloaded (Bridges, 2009, p. 502). Such criteria are more readily reducible to quantitative measures and thus constitutes one of their greatest appeal in terms of the production of some kind of evidence of research quality. As a result, research assessment exercises in AHSS disciplines now place a greater emphasis on measuring academic performance in terms of quantitative research metrics (such as: *bibliometrics, impact factors* and/ or *citation indices*), in an attempt to demonstrate greater accountability and value-for-money within this age of austerity. Ultimately there is a danger "*that measure rather than quality becomes the target of research activities, and maximising impact factor an end in itself*" (Smeyers and Burbules, 2011, p.13), or that research assessment exercises engage in "*counting only what can be counted, and of assuming that quantitative measures are inherently more objective..." (Smeyers & Burbules, 2001, p.6)*

The practice of employing quantitative metrics as a measure of quality "*obviates the need for more difficult and potentially controversial judgments, such as actually reading and arguing over the value of scholarship" (Smeyers & Burbules, 2001, p.6)*

and thus has the potential to impact negatively on the *actual* quality and diversity of research, as well as on the integrity and *'the autonomy of researchers*' (RIA, 2009, p2) undertaking AHSS research in Ireland. The discussion that ensues critically examines the assessment of research performance, and critiques some of the emerging and emergent anxieties in relation to researcher performance in assessment exercises.

Context for development of Research Metrics

The relatively recent discussions on a set of metrics suitable for assessing AHSS research in Ireland have been advanced by the work of a number of groups, particularly the Royal Irish Academy (RIA), who produced a report '*Advancing HSS Research in Ireland'* in 2006. This report recommended the development of indicators to assess the quality of Humanities and Social Science research in Ireland. Subsequently, the RIA convened a meeting of humanities scholars in 2009 in conjunstion with the Irish Research Council for Humanities and Social Sciences, IRCHSS, to discuss the development of key performance indicators "*sensitive to the unique characteristics, strengths and contributions of humanities research in Ireland"* (RIA, 2009, p2). Other Irish organisations who contributed to the development of indicators for assessing AHSS research include: the Irish Higher Education Authority (HEA) and IRCHSS, who jointly undertook a foresight exercise into the Arts, Humanities and Social Sciences in 2007 to identify "*effective metrics of research output (both quantity and quality) consistent with national and international trends*" (Barkhoff, 2008) and the Irish Universities Association (IUA), who engaged in a government funded project to develop key performance indicators for the university sector in 2008. The IUA funded project identified 32 quantitative indicators to evaluate performance of higher education institutions, with 5 of the indicators relating to research outputs, which included: PhD awards, publications in peer reviewed journals, citation in journals, research journals and international research honours.

The rationale for the development of key performance indicators to assess the quality of AHSS research in Ireland and elsewhere generally includes the need to benchmark research with national and international research; to measure productivity and value-for-money of humanities researchers who are accessing public funds; and to inform a research performance-based funding system. In 2009, the RIA President, Prof. Nicholas Canny, identified three reasons why it was important to engage Irish AHSS scholars in fresh discussions on research metrics. Firstly, at that time university management and funding agencies in Ireland were moving towards a '*science-inspired system of bibliometrics that seemed entirely inappropriate for measuring research achievement in humanities disciplines and took little account of peer review'* (RIA, 2009). This was evidenced by the aforementioned IUA project, which had identified solely quantitative indicators for research assessment exercises within the Irish higher education sector. Secondly, policymakers had little understanding of the diversity of research within humanities, and the corresponding need for a variety of key performance indicators appropriate to the varied disciplines within humanities. This was evidenced by the value that was placed on on particular types of publications (mainly journal articles), and/ or the lack of recognition for others, within research metrics being utilised by some organisations and funding agencies, such as HEA, IRCHSS and IUA. Thirdly, university rankings nationally and internationally were

impacting on the reputation of disciplines within Universities, thus there was a perceived need for humanities scholars to urgently identify metrics appropriate to their discipline or they could find themselves 'ignored or marginalised' or measured by criteria that is not relevant in their research domain. Morrissey (2013, p.807) notes how "the danger is that the emergent performance measurement culture will be locked into neoliberal and bureaucratic delineations of research and educational productivity—a regime of truth, in a sense, about academic performance." Furthermore, he articulates the need for active, critical representation by academe of what constitutes 'academic contribution' and 'impact', informed by the respective requirements and exigencies of the academic disciplines and their key stakeholders: educational and civic. According to the author, this "ultimately requires our taking seriously the key challenge of authoring and convincingly insisting upon the responsibilities, functions and values of contemporary higher education—within our own institutions and collectively to our broader publics." The process of identifying suitable research assessmentsfor AHSS is still on-going in 2014, and in the absence of unified agreement, the research assessment being promoted within higher education institutions in Ireland tend to mirror the predominantly quantitative key performance indicators utilised by funding agencies and higher education associations. Although it outlines a broad and inclusive array of creative research outputs within the AHSS, the recent (December 2013) HEA *Towards a Performance Evaluation Framework: Profiling Irish Higher Education* concludes overall, "arguably some of the research outputs/would best be assessed as indicators of researchers' and departments' civic engagement rather than as a measure of research productivity or quality." (HEA, 2013, p.44) This further underscores the importance for the academic community, and especially the AHSS, to identify and be clear upon suitable research assessments, to ensure that all important creative research outputs receive parity-of-esteem, and are appropriately recognised and valorised, alongside biblometrics and other more quantitative citation measures, e.g. h-index.

Components of Research Performance

Research performance in higher education is typically assessed through an examination of evidence of research activity, research impact and/ or research quality, and, in some cases, assessing the environment in which the research takes place.

The measurement of *research activity* usually refers to the practice of bibliometrics, which is a count of the numbers of books, chapters, journal articles, performances etc that have been published within a specific time-period. The measurement of the *impact of research* generally involves ranking the book or journal publisher and reviewing the publication's impact factor, and/ or how many times other published authors have cited your work (citation analysis).

Bibliometric analysis is usually conducted on databases that exist primarily for bibliographic purposes – Thomson ISI would be considered as a leader, alongside SCI (Science Citation Index), SSCI (Social Science Citation index) and AHCI (Arts and Humanities Citation Index). An alternative database of scholarly material includes Scopus (launched by Elsevier in 2003). Researchers using bibliometrics are often looking for an optimum h-index, (based on model by Hirsch, 2005); thus, a scholar with an appropriate h-index has published h papers in scholarly journals and has been cited by others at least h times.

The evaluation of the *quality of research* is usually assessed through some form of peer review. This can be at what is perceived as a deep level review in the form of articles that are peer-reviewed for *learned* journals. The quality of a journal for example is evaluated by reputation, adjustment to norms of international journal standards or benchmarking with international standards, and citation indicators i.e. journal impact factor (Fernandez-Cano & Bueno, 2002). The evaluation of quality of research may also take the form of a sample peer review, such as the Research Assessment Exercises (RAE)/ Research Excellence Framework (REF) exercises in the UK, which involves a review of samples of research work by a panel of experts in the relevant discipline.

Other forms of assessment of the quality of research may include esteem markers; ranging from number of PhD completions or externally examined PhDs, to being nominated for the Nobel Prize or a UN Chair. Finally, research performance is very much dependent on the research environment. The nature and quality of the research environment can be evaluatedby examining the number of training programmes for example on offer to post-graduate students, the level of access to quality resources (such as databases, books, equipment, human resources) and the completion rates of post-graduate students within an institution or research centre.

Challenges in Assessing Research Performance

Assessing research performance in any field can be difficult, but is particularly challenging within the AHSS due to the wide variety of research outputs in the domain. The framework of research metrics for humanities research suggested by a UK expert group in 2007 included:

> *research outputs; spend on research infrastructure and other funding of the research environment; peer-reviewed external research income (from the research councils, but also from other peer-reviewed sources, such as charitable foundations, overseas funding agencies, etc); and evaluation of the wider social, cultural and economic significance of the research process; PhD completions per research-active member of staff; esteem indicators (such as election to national bodies; membership of editorial boards; invitations to give named lectures, large lecture series etc) (Worten, 2007, pp.177-178).*

In 2009, the RIA in Ireland presented an even more diverse list of examples of what may constitute humanities research outputs which included the following:

> *Monographs, carbon-dating exercises (& reports of), journal articles, excavations (& reports of), papers in edited conference proceedings, translation activities, edited festschriften, compiling art portfolios, chapters in edited books/ collections, reporting on art conservation, textbooks, literary productions, general interest, popular science books, creation of data sets and databases, articles in online journals, interactive online editions of academic materials, major bibliographic work, review of a year's work in a discipline, commissioned creative work, creative writing, organisation of scholarly conferences, installation of an exhibition, co-ordination of research projects/*

> *teams, music composition & performance, editorship of scholarly journals, film making, research income, drama production, number of phd students, script-writing (drama and documentary), public service, documentary editing, book reviews* (RIA, 2009, p.7).

The relevance of quantitative indicators typically used in research assessment exercises is hard to see when one considers a research output such as an art installation or public performance of a poem for example.

There ar many challenges in using predominantly quantitive indicators in assessing AHSS research outputs. The RIA (2009) pointed to differences in patterns of publication for humanities scholars, when compared to natural and life sciences; noting that typically humantites scholars publish more monographs and scholarly books than journals (20-35%) compared to (45-70%) in social sciences (p.6). They further pointed to the dominance of single-authored publications among humanities scholars, representing a significant number of years scholarship; the corollary to this being that the citation window for humanities needed to be much longer than in natural sciences. This is further complicated by indirect bias towards women (who are more likely to change their surnames, which citation indices aren't sensitive to). Furthermore, the type of publications that citation analysis supports doesn't reflect the diversity of modes of publication in this field. The RIA (2009) accepted that more information is needed on "*activity and publication behaviours of scholars from the different disciplines"*(p.8). They further suggested that the gathering of such information should happen at the level of the individual discipline or sub-discipline.

In its summation of the limitations in the use of Bibliometrics to assess humanities research, the RIA (2009, p6&7) noted the difficulties in rating the relative importance or value of a scholarly journal within a particular discipline given the absence of an agreed list of scholarly journals in AHSS, and the poor-representation of non-journal research outputs and non-English language research. In terms of the latter, Archambault & Vignola Gagne (2004) provide a compelling justification for inclusion of non-English language research:

> *In many cases, the concepts and subjects covered in the SSH can be expressed and understood only in the language of the culture that is shaping them. Accordingly, SSH scholars publish somewhat more often in their own language and in journals with national distribution* (p16).

The issue with utilising bibliometrics in research assessment exercises is that, by themselves, bibliometrics are limited in terms of assessing research quality, and societal and cultural impacts (RIA, 2009). Also what is valued or counted within bibliometrics may not reflect the diversity of outputs that should or could be recognised within the field under examination. Smeyers & Burbules (2011, p.?) for instance point out that in a discipline such as Philosophy of Education it is more likely to refer to seminal work of philosophers such as Plato, Kant, Derrida, thus making minimal use of citation of recent articles. As a result, journals in this subject domain generate low impact factors, which may eventually lead to their demise. Smeyers and Burbules (2011) also highlight how some publications become standard reference points within certain disciplines, thus producing a snowball citation effect for the author/s, where an element of "the rich getting richer" exists.

Fernandez-Cano & Bueno (2002) note that the impact factor highlighted within Journal Citation Reports (JCR) has become the criterion for establishing the value of a researcher's work, despite heated criticism of this within the AHSS. According to Konkiel (2013) impact factors have been criticised on two main grounds: 1) gaming and 2) granularity. Gaming includes: cosmetic citations (self-citing, citation clubs among cliques of researchers); editorial boards requiring authors to cite articles previously published in their journal inflate the citation count, and inaccuracy of citations or referencing. Granularity refers to the capacity of citation indexes to provide only an approximation of the true quality of an article.

Furthermore, the premise underpinning the use of impact factors and citation analysis is that scholarly work of good quality will be frequently cited by others. However, articles may have been frequently cited to point at flaws and inaccuracies within the research, thus indicating that frequency of citation is not necessarily an indicator of good quality scholarly work. Smeyers & Burbules (2011, p.4) argue that "*inferring importance because something is cited, even cited frequently, is a leap of logic*". For instance Konkiel (2013) refers to a salient example ("A Bacterium That Can Grow by Using Arsenic Instead of Phosphorus") published in the June 2011 issue *of Science,* where nearly all citations received were from scientists disputing the hypothesis of the original article. Despite attempts made by Framework 7 EERQI (European Educational Research Quality Indicators) to explore the possibility of identifying the semantic features of 'negative citation', the project only produced limited data which might assist rather than replace a more traditional form of assessment based on professional judgment. Thus, this reinforces that negative citation is difficult to isolate and exclude from citation calculations.

Finally, even the various peer review processes have limitations. Some criticisms of peer review identified by the British Academy (2007) are that it is '*unreliable, irreproducible, biased through self-similarity or sheer greed, sexist, ignorant, careless, dishonest*' (p. 21). Deep peer review is usually associated with a particular type of publication (mainly, the assessment of scholarly journal articles), and isn't utilised in the assessment of the broader range of research outputs that may be output in AHSS. Deep peer review can slow up the publication process, allow for irresponsible behaviour or conflicts of interest on the part of the reviewer, and/ or can inhibit innovation in published work. Sample peer review, usually associated with RAE exercises, can be time-consuming and expensive (particularly where panels or boards are formed to review research), result in reviewer fatigue (RIA, 2009) and for countries like Ireland (with small academic communities) could be difficult to implement. Peer review of funding applications can be problematic as decisions are based on *prospective* quality of research (as opposed to *actual* quality, which can only be determined by retrospective judgements on the success of research undertaken), process is more subjective (as identity of applicants is not always kept anonymous) and applications generally can not be re-cycled if unsuccessful (as may be the case with journal articles). There are issues around the provision of training for peer review. The professional practice of reviewing and ethical conventions for peer review (particularly with regard to fairness in dealing with work of others) need to be more systematically examined and training in this regard needs to be provided. The composition of peer review panels and boards (in terms of wider representation from outside the academic community) also needs to be examined. Finally, according to

the British Academy (ibid, p.27) there is a need to incentivise sample peer review, as there is little or no monetary or academic recognition for engaging in a review process. It is perceived as an academic 'duty', as an opportunity to keep up with new developments or to shape the field of research and / or to gain knowledge on how funding is awarded.

International Experiences of AHSS research assessments

In a 2008 study exploring the feasibility of establishing a common approach to evaluating the outputs and outcomes of humanities research in Europe, the Higher Education Research Authority, HERA, (as reported by Barkhoff, 2008) came to the conclusion that there was poor coverage of humanities publications within internationally recognised scientific citation indexes; monographs for example were not included by the Institute of Scientific Information index even though they were a dominant mode of publication within humanities. HERA also pointed to the slower impact of publications in the humanities, thus the citation window was substantially different, with seminal or radical works taking a long time to have an impact. They pointed to the fact that non-English publications were being disadvantaged, with research in some disciplines that shared a national or regional focus, or offered inter-disciplinarity perspectives, lacking visibility or recognition at an international level. HERA highlighted that esteem indicators needed to be included alongside other research outputs, and that peer review was needed in assessing humanities research. They also pointed to the need to measure wider social, cultural and economic impacts, which necessitated stakeholder analysis, and review of performances and exhibitions, presence in media, involvement in policy-making and/ oradvisory roles for governments and business. Finally, they called for the development of more comprehensive metrics for Open Access publications.

In its review of research assessments for AHSS across the UK, Netherlands and Australia, the Royal Irish Academy (RIA, 2009, p4-5) further commented on the need for participation and leadership from within the humanities community in the identification of appropriate key performance indicators on a discipline-by-discipline basis. They warned of the need to guard against confusion between measures of impact and measures of quality, and the need to resolve the tension between assessing scholarly quality and the societal and cultural relevance of research. The RIA also highlighted differences in how the outcomes of research assessment exercises were being reported. Data for research assessment is typically collected at the level of the individual scholar but collated and reported at the level of the School in the UK, at the level of Research group or centre in the Netherlands, and at the level of discipline or sub-discipline in Australia. Finally, they pointed to problems in identifying indicators for early career researcher, in interdisciplinary fields, or of teaching and learning activities.

It is interesting to note that in the former UK RAE exercises, the Education panel judged submissions on their own merits rather than on the pedigree (place or form) of publication. According to Bridges (2009), this practice was enacted to counter the absence of an agreed hierarchy of publications "*which could serve as a proxy for a more direct judgement of quality"* (p.504) in educational research, to address the inconsistency among educational journals in maintaining high standards of quality,

and furthermore to recognise that submissions published in locations other than high ranked journals can also be of a high quality. This 'open-minded' practice would be beneficial in capturing the value of quality publications within research fields such as Irish Studies, which are of interest to a very small community of research scholars mainly based in Ireland, and whose publications tend to be local or national with limited international visibility. A further concern raised by Labaree (2008) is the ill-judged practice of attempting to make educational research more relevant to what is being valued by authorities at a particular point in time; stating that valuing one form of educational research over another (for example valuing practice based or applied research over theoretical research) does not make sense since research tied to a particular context may 'grow stale' very quickly, whereas more theoretical or scholarly based research may prove to 'age well' (p.423). Key performance indicators of research within the AHSS must be sensitive to quality research within such domains, rather than re-shaping research outputs in sub-disciplines of AHSS like education to conform to external agendas or assessment exercises.

In the RIA's (2009) four-part typology of research impacts that emerged from discussions with AHSS scholars in Ireland, it is argued that the impact of research can be assessed according to whether its impact and / or contribution are to "*a) promote academic excellence and impact; b) enrich the scholarly community; c) encourage teaching and learning; and d) contribute to civil society*" (RIA, 2009, p.8). Furthermore, the RIA (2010) report that humanities scholars recognise that metrics-informed indicators have a role to play in assessing quantitative items such as research income, number of PhD students, and that appropriate bibliometrics can add '*empirical content and objectivity to the process of evaluating research outputs'* (p.9). The RIA also noted that the appropriate *unit of measurement* for research assessment exercises should be at the level of the discipline, research cluster or institute within the institution (rather than at the level of the individual researcher). They also stated that research assessment frameworks must be sensitive to the differences between disciplines and sub-disciplines and allow for discipline variation. The RIA consultation phase with humanities scholars concuded that peer review was a critical dimension of any research assessment for the humanities in Ireland, noting that '*there is a general consensus [among humanities researchers] that peer review must be at the heart of research assessment for the humanities*.' (RIA, 2009, p.2)

Finally, Vella (2013, p1) comments that the '*problem faced by creative artist researchers is the privileging of one approach to explanation and plausibility over others'*. Many research assessment frameworks (such as the Australian Excellence Research Australia) differentiate between research outputs as being 'traditional' or 'non-traditional', with the former inevitably being more valued in research metrics exercises. As those among us in academia are well aware, non-traditional is generally anything that falls outside of the categories of journal articles, books, chapters and manuscripts. Vella (2013) argues that the latter labelling results in research being "defined by being *not* rather than being *something*" (p1); such as *being a research output* which in itself demonstrates or presents multimedia elements, artistic representations (sculpture, imagery) and/ or performance (dance, musical) etc. The key tensions in the lack of recognition or value being apportioned to these non-traditional research outputs in the research metrics exercises is indicative of three issues – "transferability, plausibility and the tools of representation" (Vella, 2013, p4). Vella (2013, p1) believes that the creative arts researcher must make the case for the

'artistic endeavour' constituting research, otherwise his/ her work simply may become fodder for the university publicity machine – in other words relegated to becoming a good news story. To make a case for recognition of non-traditional research, the expectations are that the artistic work is contextualized and its knowledge claims explicated by the researcher. These are processes that researchers in any discipline can readily engage with; however even in doing so, the research work can be still be relegated because of preferred modes of representation of research (and thus political dispositions towards research) within and across disciplines/ faculties/ institutions.

Research Assessment of Educational Research in Ireland

In Ireland, the Research Assessment in Education working group was established in March 2009, as those within the educational research community in Ireland became aware that problematic conflations could arise if clear and relevant criteria were not identified by scholars themselves. The goals of this working group were to encourage free and open discussion on the application of research metrics within educational research, to examine the diversity of research outputs within the discipline of educational research and to reach consensus on how best to progress towards identifying a framework of assessing the quality of research within the discipline of educational research in Ireland. A series of six meetings were convened from March to December 2009, with contributions from four of the seven universities, the National College of Art and Design, three colleges of education in the south of Ireland, and one university from the north of Ireland. Representation at these meetings varied from heads of departments, research convenors or other senior persons nominated by the institutions.

The outcome of the working group was to recommend a 5-level research profiling system that captured an integrated range of work at the level of a unit (school or department) rather than at the level of the individual scholar. The five levels describe a continuum from a research focused department of high standing, with a wide diversity of research activity, to one with little or no research focus. The profiling system focuses on research outputs, with a sample list of research outputs and esteem indicators outlined in an accompanying document. A principal objective of this model was to differentiate between research and non-research work while also avoiding defining 'non-research work' negatively. While acknowledging that there are journals, publishers and funding agents of high esteem, the system takes a qualitative line that is sensitive to varying size and scope of what could be considered research activity within individual education departments in Ireland.

The 5-level research profiling system was utilised as a basis for discussion on suitable research metrics in education with colleagues within Irish universities and colleges of education. It was also presented at a symposium by Holland and Hall (2010) at the Educational Studies Association of Ireland Conference in March 2010. This facilitated critique, feedback and input from colleagues across the island of Ireland, supporting the further development of the valorisation framework and related esteem and other indicators for educational research in Ireland. A number of clarifications, suggestions and amendments emerged through this process. Firstly, there was recognition that there was a need for some kind of assessment framework for assessing quality of educational research in Ireland, and that the 5-level profiling

system and accompanying list of research outputs and esteem indicators were useful in that sense – a necessary 'evil' was the view expressed by one participant. Criticism of the five-level model included that the levels needed to be more clearly defined and the model was limiting in terms of what was valued at level 1 in particular. The issue of what was valued was compounded by the 'hierarchy' of research outputs in the research outputs document. There was agreement that the research outputs should not be listed as 'in order of prestige' (rather there should be equal weighting for each item) and the list should not be considered a definitive list of research outputs (edited works for example didn't appear on this list). The removal of any hierarchy of research outputs is supported by the RIA (2009) in their comments that "*research assessment should seek to measure the quality of specific outputs rather than attempting to create a false hierarchy of output types, particularly in respect of publications*" (p8).

There was consensus that civic engagement should feature on the list of research outputs. Furthermore, it was noted that impact in terms of social and cultural impacts could be measured with the inclusion of 'case studies' that described a societal or cultural impact as was the case in the new REF system in UK. Finally, there was discussion on how the research 'environment' would be valued – this was deemed very important in terms of supporting the research process.

Finally, the 5-level research profiling system is based on the premise of both sample and deep peer reviews. The peer review process is important in indicating the quality of research. However, one needs to be careful need to be careful about rushing to value impact (social etc) over quality; thus, one needs to avoid valuing what is topical in terms of general public interest over quality of research, and there must be rigour in the research process. Peer review is important in terms of identifying malpractice (presentation of false findings), plagiarism, redundant publications (presenting same material multiple times) and breaches in research integrity such as 'Salami-slicing' (presenting more publications than is reasonable for a single study).

Future Metrics: The Rise of Altmetrics

The many limitations of traditional measures of research quality and impact, such as journal impact factors and article citations, have given rise to much debate, and the call for metrics that give an appropriate measure of research quality and impact. Google Scholar (launched by Google in 2004) is a search engine that has been optimized to isolate scholarly publications from online databases and provide citation statistics. Google Scholar searches for book and chapter publications, unlike the ISI index, and captures these citation outcomes in the realization of impact factors. Furthermore, Google Scholar has the potential to capture citation and reference of scholarly works embedded in graduate dissertation, theses, and lecture notes that previously have not been utilized in citation counts and impact factors. Google Scholar uses a variety of methods and algorithms to systematically search the Web for intra-citations and inter-citations of a range of articles, chapters, books and other scholarly sources on the Web, that are ranked according to spheres of 'influence'. This differs from Scopus and Web of Science in that Google Scholar not only counts citations but it bases their analysis on the importance of the article within the scholarly context. This has the added benefit of isolating contributions that have had a significant impact on a particular discipline or field. However, Van Aalst (2010)

comments that '*Citation counts obtained from Google Scholar may exaggerate impact, and the citing documents may not be scholarly or peer reviewed'* (p. 387).

The emergence of Altmetrics, alternative metrics, a term coined by Jason Priem in 2010, refers to the online digital footprint of research or the impact it has had within the online context. It represents the latest move towards integrating multiple metrics in the realization of an impact factor for research outputs, particularly the translational impact of research online. According to Howard (2013), the emergence of Altmetrics have now made it possible to track and share evidence of the impact of research online, by using new facilities that collect data (number of hits/ views, tweets, likes, citation counts, downloads) from social media (blogs, twitter, discussion forums, online newspapers, Figshare -sharing research data-, slide share etc) reporting on research, as well as from more traditional online research documents (books, chapters, articles, reports). Different levels of data are given different weighting (tweet weighted more than a like in Facebook). Some publishers have already article-level metrics – tracking articles published by the publisher. In his report in August 2012 on Altmetrics, Paul Jump noted that researchers unhappy with "crude journal metrics" are already turning to alternative metrics that show the real impact of their work. He further highlighted that Zotero, ResearchGate, CitULike, BibSonomy and Mandeley are suited to the generation of alternative metrics because their primary function is to allow researchers to share and engage in discussion on research.

Key anxieties in assessing research performance

Research assessment exercises are now also used to determine who gets grants and infrastructural funds from government, research councils and other bodies within and beyond Europe. Within these assessment exercises, governments and funding agencies are becoming reliant on more 'empirical' or 'numerical' ways of assessing the research outputs of universities. In Spain, as a result of decrees by parliament from 1989 to 1996, Spanish researchers' productivity has been mainly assessed on the publication of research articles in international ISI-indexed journals with a high impact factor. According to Delgado López-Cózar, Ruiz-Pérez & Jiménez (2007), the impact of these policies has led to a 255% increase of Spanish research articles per year in the ISI database, and the internationalization of Spanish research, with corresponding improvements in the rigour, quality and impact of research. However, the mass emigration of the best Spanish research articles to mainly English language international journals has also resulted in the neglect of Spanish journals (which often have to 'make do' with the publication of research of lesser value), and more worryingly '*the destruction of Spanish as a language of science'* (Delgado López-Cózar, Ruiz-Pérez & Jiménez, 2007). Furthermore, Delgado López-Cózar, Ruiz-Pérez & Jiménez, 2007 note there is evidence of a move away from research with a local, regional or national value within Spain, to more generic research that is valued by the international community. Finally, Delgado López-Cózar, Ruiz-Pérez & Jiménez (2007) report a corresponding rise in 'impactitis' (coined by Cami, 1997), which Delgado López-Cózar, Ruiz-Pérez & Jiménez (2007) elucidate as "*altered publication and citation behaviour in response to an obsessive compulsion to use the impact factor as the single, incontrovertible quality criterion for scientific articles" (ibid),* which they report as a disease of epidemic proportions in Spain evidenced by

the extent of self-citing, citation clubs and publication decisions based on the impact factor of journal alone rather than on the most appropriate audience for work.

According to RIA(2009) the growing trend in Europe to consider the humanities and social sciences as a single field in research policy will inevitably result in research agendas and key performance metrics from the natural and life sciences being uniformly applied to the humanities – *"Such indicators will not capture the impact and contributions of humanities research*" (p.3). In this respect, Bridges (2009) argues that we should not underestimate the challenges in assessing the quality of educational research, highlighting challenges posed by the *"overwhelming diversity in the theoretical framing of educational research and in the epistemological and ontological assumptions which underlie it, let alone the methods employed and the forms in which it is subsequently represented"* (p.499). Furthermore, ISI impact factors are low in applied AHSS fields like Education. According to Val Aalst (2010: 388), "*the median ISI impact factor of all journals in the Education and Education Research category was 0.548"* in 2007; this can lead to a perception of educational research being of a lesser value than research within other disciplines, or educational researchers within specialized areas being less competent than their peers. Another issue with ISI is the length of time taken to publish in higher ranked journals is much longer in education than other disciplines, which according to Van Aalst (2010: 388) delays citations and ultimately results in lower impact factors. Therefore, there is an urgent need for AHSS scholars to reach consensus on ways of measuring the quality of their research. Research quality assessment will shape researcher behaviour mainly because, as Bridges (2009) outlines "*what counts as good research subsumes a set of principles about what will count as research at all*" (p498). Thus, if journal publications are prioritised as in the aforementioned case of Spain for example, then this will likely be followed by a move away from book publications and over a period of time such inclusions or exclusions will affect what knowledge and in turn what research is valued in the university. The process of determining what counts as quality needs to be carefully managed. As Bridges (2009) noted, the assessment of research quality is one of the primary drivers of the behaviour of the academic community with all sorts of potential for distorting that behaviour.

Finally new modes of publication, such as online journals, social media outputs, need to be critically examined for inclusion in future research assessment exercises. The concerns around altmetrics include a lack of knowledge on who controls the sources of data online, resistance among techno-phobic academics to engage in altmetrics, and the heightened risks of gaming and vulnerability to corruption in citation analysis. In the case of the latter, this refers to the manipulation of *meta-data* (key words that describe the research) that has the potential to unfairly increase the citation indices of those researchers with a better understanding of *search optimisation techniques* within online contexts. Furthermore Konkiel (2013) argues that altmetrics providers have yet to develop a way to differentiate between scholarly and sexy research (research that is topical and popular beyond academic circles) and she contends that, as in the case of other forms of metrics, it "*does not apply as readily to traditional works such as books or art*" (p.15).

Conclusions

AHSS researchers are becoming increasingly concerned at the lack of appropriate metrics and inclusivity in what areconsidered valued research outputs. Evidence is emerging that the mainly quantitative metrics system is changing the behaviour of researchers towards research that can be counted (Nature, 2010) and in some cases being used at an institutional level to wean out non-research active staff for redundancy or teaching-only positions. Early stage researchers need guidance early-on on how best to progress, as performance anxiety increases in the absence of a clear and cogent criteria and guidance on what should be valued in AHSS research. There is a need to consider *good* international practice in measuring research quality, whilst also recognising that key performance indicators developed elsewhere need to be customised or adapted within the context of AHSS research in Ireland (RIA, 2009, p3). There is a need to advocate recognition of the diversity of research within applied AHSS disciplines, such as education, and to recognise that while metrics are of value, peer review is of equal value (Wooten, 2007) and in some cases considered indispensable (Barkhoff, 2008) in quality research assessment exercises in AHSS.

Acknowledgements
The authors would like to acknowledge and thank members of the working group on Research Metrics in Education, and those that have contributed feedback in the process.

REFERENCES

Altbach, P.G. (2006) The Tyranny of Citations. Inside Higher Ed. Available at: http://www.insidehighereducation.com/layout/set/print/views/2006/05/08/altbach Accessed on 13/04/2010.

Archambault, E. & Vignola Gagne, E. (2004) The Use of Bibliometrics in the Social Sciences and Humanities. Prepared for the Social Sciences and Humanities Research Council of Canada (SSHRCC) by Science-Metrix. August 2004

Barkhoff, J (2008) Research Evaluation, Metrics and KPI's in the Humanities: The situation in Ireland and its European context. [Online] Workshop in Trinity College Dublin in September 2008 jointly organised by Coimbra Group Task Force Culture, Arts and Humanities and HSIS Humanities Serving Irish Society Consortium. Available at: <??. Accessed…

Benneworth, P & Jongbloed, B. (2009) Who matters to Universities? A Stakeholder perspective on humanities, arts and social sciences valorisation. High Educ (2010) 59:567-588. DOI 10.1007/s10734-009-9265-2

Bridges, D. (2009) Research quality assessment in education: impossible science, possible art? British Educational Research Journal. Vol. 35. No. 4, August 2009. PP. 497-517.

British Academy (2007) Peer Review: the challenges for the humanities and social sciences. Sept 2007. Available at: http://www.britac.ac.uk/

Boletín Oficial del Estado, Pub. L. No. 28563, (September 9, 1989). Boletín Oficial del Estado, Pub. L. No. 3566, (February 6, 1990). Boletín Oficial del Estado, Pub. L. No. 37030, (December 3, 1994). Boletín Oficial del Estado, Pub. L. No. 35028, (November 20, 1996).

Camí J. Impactolatría: diagnóstico y tratamiento. Med Clín (Barc) 1997; 109: 515-24.
ERA (2010) ERA 2010 Submission Guidelines. Excellence in Research for Australia. Research in the national interest – enabling the future. Commonwealth of Australia 2009. ISBN 978-0-9806204-5-0.

Corbyn, Z. (2010) DIUS abandons plans for different REF systems to judge sciences and arts. [online] Times Higher Education. Available at: http://www.timeshighereducation.co.uk/story.asp?storycode=401577 Accessed: 14/04/2010

Fernandez-Cano, A. & Bueno, A. 2002 *Multivariate evaluation of Spanish educational research journals*. Scientometrics, Jointly published by Akadémiai Kiadó, Budapest and Kluwer Academic Publishers, Dordrecht. Vol. 55, No. 1 (2002) 87–102

Gibney, E. (2013) *Publication won't lead academy out of technological death valley*. Times Higher Education. 21st March 2013

Higgins, M. D. (2013) Toward an ethical economy, Ethics for AllPublic Lecture Series, Dublin City University, 11th September 2013 Available at: http://www.president.ie/speeches/toward-an-ethical-economy-michael-d-higgins-dublin-city-university-11th-september-2013/Accessed: 14/11/2013

Higher Education Authority of Ireland. (2013) Towards a Performance Evaluation Framework: Profiling Irish Higher Education Available at: http://www.hea.ie/sites/default/files/evaluation_framework_long.pdf Accessed: 20/02/2014

Holland, C. and Hall, T. (2010) Symposium on Research Metrics in Education. Educational Studies Association of Ireland 35th Annual Conference: Borders, Boundaries and Identities: Education in Challenging Times. 25th-27th March, Dundalk.

Howard, J. (2013) Rise of 'Altmetrics' Revives Questions About How to Measure Impact of Research. The Chronicle of Higher Education: Technology. June 3rd 2013.

Jump, P. Research Intelligence - Alt-metrics: fairer, faster impact data? Times Higher Education. 23 August 2012.

Jump, P. (2012) Research intelligence - Value according to the metrics system. Times Higher Education. 26th April 2012.

Konkiel, S. (2013) Almetrics. A 21st-Century solution to determining research quality. Online Searcher, Jul-Aug 2013 pp. 11-15

Labaree, D.F. (2008) Comments on Bulterman-Bos: The Dysfunctional Pursuit of Relevance in Education Research. Educational Researcher. 37:7, 421-423.

Morrissey, J. (2013) Governing the academic subject: Foucault, governmentality and the performing university, Oxford Review of Education, 39:6, 797-810, DOI: 10.1080/03054985.2013.860891.

NATURE (2010) Do metrics matter? Vol. 466, pp.860-862, 17th June 2010.

Nussbaum, M. C. (1990) *Love's knowledge: essays on philosophy and literature* Oxford, Oxford University Press.

Pendlebury, D A. (2008) White Paper Using Bibliometrics in evaluating research. Thomson Reuters.[A13}

Research Information Network & JISC (2009) Communicating knowledge: How and Why UK researchers publish and disseminate their findings. A Research Information Network report. September 2009. http://www.rin.ac.uk/communicating-knowledge

Royal Irish Academy, RIA (2009) Developing Key Performance Indicators for the Humanities. A report of a meeting convened by the Royal Irish Academy and the Irish

Research Council for the Humanities and Social Sciences. 12th March 2009. Royal Irish Academy.

Smeyers, P. & Burbules N. C. (2011) How to Improve your Impact Factor: Questioning the Quantification of Academic Quality. Journal of Philosophy of Education . 45:1 pp. 1-17

Smith, R. (2010) Poststructuralism, Postmodernism and Education IN Bailey, R., Barrow, R. Carr, D, and McCarthy C. (Eds) The Sage Handbook of Philosophy of Education, pp. 139-150, Sage, London

The Guardian: Higher Education Network. [Online Forum] Twitter, peer review and altmetrics: the future of research impact assessment. 21 September 2013.

Van Aalst, J. (2010) Using Google Scholar to Estimate the Impact of Journal Articles in Education. Educational Researcher. 39:5 pp.387-400.

Vella, R. (2013) The Rhapsode goes to University. A discussion of Plato's Ion in relation to creative arts research metrics.

Worton, M. (2007) Of models and metrics: the UK debate on assessing Humanities research. Peer Review, its present and future state. Prague for Science 2006, 12-13 October 2006.