

The Multi-Dimensional Research Agendas Inventory – Revised (MDRAI-R): factors shaping researchers’ research agendas in all fields of knowledge

Hugo Horta^{1,*} and João M. Santos²

1 Division of Policy, Administration and Social Sciences Education, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong, SAR, China

2 Instituto Universitário de Lisboa (ISCTE-IUL), Centro de Investigação e Intervenção Social (CIS-IUL), Lisbon, Portugal

*corresponding author: horta@hku.hk

Hugo Horta ORCID: 0000-0001-6814-1393

João M. Santos ORCID: 0000-0001-6414-6891

Editor: Ludo Waltman

Date received: May 6, 2019

Date accepted: November 24, 2019

Abstract

This study creates a novel inventory that characterizes factors influencing the research agendas of researchers in all fields of knowledge: the Multi-dimensional Research Agendas Inventory-Revised (MDRAI-R). The MDRAI-R optimizes an initial inventory designed for the social sciences (the MDRAI) by reducing the number of items per dimension, improving the inventory’s psychometric properties, and including new dimensions (“Academia Driven” and “Society Driven”) that reflect the greater influence of social and organizational structures on knowledge production and demands for research impact. This inventory enhances our ability to measure research activities at a time when researchers’ choices matter more than ever, and will be of interest to researchers, policy makers, research funding agencies, and university and research organizations.

Keywords: Research agendas; Survey instrument; Instrument validation; Research choices; Research management

Author roles:

Hugo Horta: Conceptualization, Funding acquisition, Investigation, Project Administration, Writing – original draft, Writing – review & editing

Joao M. Santos: Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing

Copyright: © 2019 Hugo Horta and João M. Santos. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

37 **Competing interests**

38 The authors declare that they have no competing interests.

39 **Data availability**

40 The authors declare that the data used in the analysis reported in this manuscript cannot be
41 made publicly available due to privacy concerns, as the informed consent form signed by the
42 participants stipulated that the data would only be available to the authors.

43 **Acknowledgments**

44 This work was supported by doctoral grant PD/BD/113999/2015 from the Fundação para a
45 Ciência e Tecnologia, and co-funded by the European Social Fund and the Portuguese Ministry
46 of Science and Education. This research was also funded by the Research Grants Council (Hong
47 Kong) through a project entitled “Characterizing researchers’ research agenda-setting: an
48 international perspective across fields of knowledge” (project number: 27608516). The authors
49 contributed equally to this work.

50

51

52

1. Introduction

With research playing an increasingly central role in driving knowledge creation in fast-paced, globalized, connected, uncertain, and technology driven contemporary societies, it is critical to better understand the factors that influence researchers' research agendas, particularly those based in academic settings. This is important not only for researchers, but also for those looking to create added value from the available research, such as policy makers, research funding agency managers, and university and research laboratory administrators (Ciarli & Ràfols, 2018; Franzoni & Rossi-Lamastra, 2017; Wallace & Ràfols, 2018). Understanding the factors that influence researchers' research agendas is ultimately relevant to the development of science itself at a time when researchers are facing global, multifaceted, and increasing complex challenges, and more and more research output is being produced without necessarily leading to breakthroughs (Young, 2015). Today, a key premise in science is that researchers' strategic research choices matter, because these choices (which are to some extent personal in nature) shape the knowledge produced and the general orientation of the broader research efforts and future research directions (Polanyi, 2012). Although researchers' choices of research agendas have been examined in seminal works in the sociology of science (Zuckerman, 1978), the area remains underexplored and has mostly been analyzed from a qualitative perspective (Luukkonen & Thomas, 2016; McGrath, 1981; Shwed & Bearman, 2010).

The literature shows that the cultures, traditions, and dispositions of fields of knowledge have a fundamental influence on researchers' choices of research (Becher & Trowler, 2001). Disciplinary cultures become embedded in the *habitus* of researchers, as they feel that they belong to and identify with specific knowledge-based research communities and abide by these communities' values, norms, and attitudes (Bourdieu, 1975). This occurs as part of a path dependent process that begins with the researchers' socialization through their doctoral studies to become independent researchers (Jung, 2018; Mantai, 2017). During this time, the researchers learn how to conduct research while accumulating expertise and developing, under supervisory guidance, research interests that are likely to resonate with and influence the current and future research choices (Åkerlind & McAlpine, 2017; Brew, Boud, & Malfroy, 2017). Research agendas can be influenced by students' mentors during their doctoral studies and in the years after completion. Collaboration with peers and other stakeholders can also influence the design of research agendas, as collaborations bring novel information, expertise, and perspectives, and the possibility of serendipitous opportunities to engage in innovative, disciplinary, and multidisciplinary research (Kingdon, 2013; Shi, Foster, & Evans, 2015).

86 The patterns of collaboration are increasingly likely to influence the research agendas of
87 researchers at a time when their career trajectories are increasingly non-linear (Hancock &
88 Walsh, 2016). Nonetheless, prestige and recognition by peers in the field continue to be critical
89 signals of important contributions to the pool of knowledge and tend to drive successful careers
90 (Kim & Kim, 2017). In the “publish or perish” research environment, where performativity has
91 become central to career survival and progression, researchers might well be encouraged to
92 engage in research agendas that promise prolific research output (i.e., publications) with high
93 levels of visibility and recognition (i.e., citations), and possibilities of funding. According to the
94 Mertonian rationales of science (e.g., the Matthew effect and cumulative advantage in science),
95 such output can lead to further publications, visibility, funding, and collaboration, including
96 invitations to collaborate in others’ research agendas (Allison, Long, & Krauze, 1982; Merton,
97 1968). These activities and dynamics define and are defined by the research agendas of
98 researchers through interactive processes, as researchers position themselves (and their
99 interests) within their research communities (Whitley, 2000).

100 A few recent studies add to our understanding of researchers’ choices of research agendas from
101 a quantitative perspective (Foster, Rzhetsky, & Evans, 2015; Horta & Santos, 2016; Santos &
102 Horta, 2018; Ying, Venkatramanan, & Chiu, 2015). These studies mainly focus on a single field of
103 knowledge or disciplinary area, such as biomedicine (Foster et al., 2015) or higher education
104 (Santos & Horta, 2018). Interestingly, these studies examine the tensions between the two main
105 research strategies identified by Kuhn (2012), that is, between the conservative research
106 strategies that are part of “normal science,” and are characterized as safe and representing
107 incremental contributions over time, and riskier strategies that tend to be more innovative and
108 disruptive in searching for new paradigms. Only one of these quantitative studies offers an
109 inventory for identifying the factors that influence the research agendas of researchers (Horta
110 & Santos, 2016). Although, to the best of our knowledge, this inventory, which is termed the
111 “multi-dimensional research agenda inventory” (MDRAI), is the first of its kind, it was designed
112 with social science researchers in mind. Our study aims to extend the MDRAI. Using a dataset
113 on over 12,000 researchers located all over the world and from all fields of knowledge who
114 provided key information about their research agendas in an online survey carried out in 2017
115 and 2018, we develop a novel instrument that identifies the key factors influencing the research
116 agendas of researchers in all fields of knowledge. Our MDRAI-R optimizes the initial MDRAI
117 developed by Horta and Santos (2016) by reducing the number of items in each dimension of
118 the original inventory, and including new dimensions relevant to fields of knowledge not

considered in the original instrument. Moreover, our revised MDRAI-R is valid for all fields of knowledge.

This study largely focuses on the methodological development of the MDRAI-R. To a lesser extent, it also stresses, wherever applicable, the substantive insights that underline its evaluative applicability in current knowledge producing settings. The methodological development of the MDRAI-R is based on a pilot study and a comprehensive psychometric evaluation that includes exploratory factor analysis, confirmatory factor analysis, validity, reliability, and sensibility evaluations, and tests of measurement invariance.

2. From MDRAI to MDRAI-R

The MDRAI is based on the classical tenets of the sociology of science and focuses on researchers' personal and environmentally influenced motivations. It is also based on the literature on academic research and work, and the changing world of science, research, and academia that underlines the increasing importance of networking, competitiveness, and resources (Horta & Santos, 2016). The MDRAI covers eight dimensions, four of which have sub-dimensions. The first dimension is Scientific Ambition, which refers to the desire for recognition by peers, as most researchers strive to have their contributions to knowledge acknowledged by their peers and gain prestige by doing so (Latour & Woolgar, 2013). This dimension has two sub-dimensions. Prestige, which represents the researcher's desire for recognition and Drive to Publish, associated with the need to produce concrete evidence of the creation of new knowledge through the proper channels recognized by the knowledge community as appropriate for disseminating and increasing the credibility and visibility of knowledge. The second dimension, Convergence, relates to the researcher's preference for research agendas that have a clear disciplinary focus. This dimension refers to a researcher's decision to build a position of authority in a sole disciplinary field. Although this usually takes a substantial amount of time (Allison et al., 1982), it can be part of a specialization strategy linked to higher research productivity gains because it avoids the transaction costs of disciplinary mobility (Leahey, 2007). Convergence has two sub-dimensions: Mastery, representing the expertise of a researcher in a given field, and Stability, the investment of time and effort in a specific discipline to become an expert in the field. The third dimension, Divergence, stands in opposition to the second dimension, as it represents the researcher's preference for research agendas that integrate or make use of more than one discipline. This dimension also has two sub-dimensions: Branching out, which refers to expanding the research agenda towards other fields of knowledge (including

the use and application of theories and methods from one field to another), and Multidisciplinarity, which is associated with the inclination to engage in multidisciplinary projects (Schut, van Paassen, Leeuwis, & Klerkx, 2014).

Discovery and Conservative, the fourth and fifth dimensions of the MDRAI, are also in opposition to each other, although these dimensions do not have sub-dimensions. Discovery refers to a researcher's preference for a research agenda that is riskier but has the potential to create new knowledge in a disruptive way, possibly creating new paradigms (Kuhn, 2012). Conservative measures the preference for pursuing a research agenda that is focused on well-established themes, and a more incremental knowledge creation perspective. This preference is deemed to be safer and within the bounds of normal science, according to Kuhn (2012), and thus entails less risk of encountering research dead-ends or a lack of acceptance by the research community. The sixth dimension, Tolerance of Low Funding, measures a researcher's willingness to pursue a research agenda even when little or no funding is available to support it. This dimension is relevant because it is associated with the competitive drive for research funding that universities and other institutions exhibit even when their researchers do not necessarily need such funding to do their research (Roumbanis, 2018). However, this dimension also illuminates how researchers can engage in research agendas without having access to resources at a time when the distribution of resources is characterized by inequality and increasing concentration (Hicks & Katz, 2011). The seventh dimension, Collaboration, plays an increasingly key role in the contemporary research dynamics (Kwiek, 2018), and refers to the preference for engaging in collaborative research agendas. This dimension also has two sub-dimensions, which represent how engagement in collaborative research can occur: Willingness to Collaborate, which indicates the propensity to collaborate, and Invitations to Collaborate, which refers to the collaborative opportunities provided by others (i.e., research projects started by others). The final dimension of the MDRAI is Mentor Influence, which measures the extent to which researchers are influenced by their mentors when designing their research agendas. The influence of a mentor on an individual's research agenda is to some extent a proxy for scientific independence, but can also attest to good professional relationships forged during a researcher's PhD study, even though the mentor's influence is expected to wane over time (Ooms, Werker, & Hopp, 2018).

The MDRAI covers these critical dimensions and can be complemented by additional dimensions that are likely to shape the way that research is thought about and considered. Based on the literature, three dimensions are considered. First, the research agendas of researchers in the fields of science, technology, mathematics, and engineering (STEM) are known to be more

Copyright: © 2019 Hugo Horta and João M. Santos. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

influenced by their field communities in which consensus on the significant questions that should be addressed tends to be reached collectively and holistically. This consensus is expected to influence a researcher's choices in those fields when defining a research agenda (Becher & Trowler, 2001). However, the research preferences of researchers in the social sciences and humanities tend to relate more strongly to personal interests. Although these personal interests are linked with issues significant to the researchers' field communities, the field communities are not expected to influence individual researchers to the same extent that they do in STEM fields (Collins, 1994). Second, with the rise of performativity, managerialism, and metrics associated with world university rankings and competitive national funding schemes, universities and other institutions are playing an ever greater role in influencing the research agendas of researchers (Kenny, 2018). These organizationally determined metrics establish the goals and targets related to research careers, and influence decisions on salary increases and tenure and promotion (Acker & Webber, 2017). The recent literature shows that the increasing institutional pressure is influencing academic work and the way that researchers use these institutional constraints and incentives to orient their intellectual interests and career trajectories (Brew, Boud, Crawford, & Lucas, 2018). Third, as research funding agencies and other institutional bodies (including universities, through policies related to research exchange) are increasingly highlighting the impact and social relevance of research, it is becoming increasingly likely that forms of research practice such as "action research communities" or "participatory research" are chosen. In these practices, researchers work collaboratively or consult lay communities about the challenges that they may face, and structure their research from this perspective (Mendes, Plaza, & Wallerstein, 2016; Wooltorton et al., 2015). As a result, researchers may increasingly seek the opinions of non-experts about social and technical problems, and build research agendas that deal with "real problems" and are likely to have a strong societal impact.

3. Method

This section provides information relevant to the various analyses presented later in this study, such as the methods of determining validity.

3.1. Structural Equation Modeling

This sub-section provides a brief introduction to structural equation modeling (SEM) to enable readers unfamiliar with this methodology to better understand the remainder of the study. Readers already familiar with SEM may wish to skip this sub-section.

In the pilot and main studies, SEM was implemented using AMOS 24, with the goal of conducting confirmatory factor analysis (CFA) as a follow-up to a previously implemented exploratory factor analysis (EFA). The AMOS software package was developed by IBM as a companion to the more well-known SPSS, focusing on SEM. Although there are other software packages dedicated to SEM, AMOS has the distinct advantage of being largely graphics-based and is thus easier to use.

SEM has the capacity to include latent variables to account for factors that cannot be directly observed (Bentler & Weeks, 1980) while also providing linear modeling procedures such as analysis of variance and linear regression (Marôco, 2010). It has also the advantage of providing significantly more fit indicators than those available for general and generalized linear modeling, which can be used to re-estimate the model to achieve optimal fit, such as by allowing for covariance between the error terms (Bollen, 2014; Maroco, 2007; Marôco, 2010).

SEM typically comprises two components: the measurement model and the structural model. The measurement model examines the trajectories from the manifest variables to the latent variables, with the dependent or endogenous variables being represented as follows (Bollen, 2014; Maroco, 2007; Marôco, 2010):

$$y = \Lambda_y \eta + \varepsilon$$

where y is the vector for the manifest variables, Λ_y is the matrix for the factorial weights of η in y , η is the vector for the latent variables, and ε is the error term for y . The independent or exogenous variables are given by:

$$x = \Lambda_x \xi + \delta$$

where x is the vector for the manifest variables, Λ_x is the matrix for the factorial weights of ξ in x , ξ is the vector for the latent variables, and δ is the error term for x .

The second component in SEM, the structural model, defines the relations between the various latent variables, and is given by the following (Bollen, 2014; Marôco, 2010):

$$\eta = B\eta + \Gamma\xi + \zeta$$

where B is the coefficient matrix for the latent endogenous variables in the structural model, Γ is the coefficient matrix for the latent exogenous variables in the structural model, and ζ is the vector for the disturbance terms in the structural model.

CFA is a specific type of SEM that is largely centered around the measurement model, because the structural section, if it exists, is largely reserved for second-order constructs. CFA is frequently used as a follow-up analysis to EFA. In EFA, the variables are allowed to freely load onto any extracted factors (Maroco, 2003), whereas CFA requires that the researcher specifies the structure to be tested (Brown, 2015). Thus, EFA can provide initial insights into how to specify the model, and this specification can subsequently be tested through CFA.

Rather than relying on ordinary least squares, various methods can be used to estimate the parameters in SEM. The de facto standard in SEM estimation is maximum-likelihood (ML) estimation. ML estimation was used in all of the SEM analyses in this study because it is robust to deviations from the multivariate normality, and is generally considered to be the most useful estimation method (Arbuckle, 2007; Jöreskog & Sörbom, 1989; Marôco, 2010).

3.2. Considerations when using SEM with a large sample

The main study used a much larger sample than is typically encountered in studies or referred to in statistical textbooks. Although this increases statistical power, it also creates issues in SEM due to the method's reliance on the χ^2 statistic. χ^2 is a mathematical function of the sample size, and is generally inflated by large samples (Hair, Black, Babin, Anderson, & Tatham, 2007). This makes the underlying test almost always significant, and other indicators that are dependent on this statistic are likewise influenced. In other words, the χ^2 statistic reflects the sample size rather than the model fit (Browne & Cudeck, 1993). As Iacobucci (2010) states, "as N increases, χ^2 blows up," with quasi-exponential gains in the χ^2 statistic reached for sample sizes as low as 500. As a result, fit evaluation was conducted using a suite of alternative fit indices (AFIs) (Barrett, 2007; Browne & Cudeck, 1993; Kline, 2016; Putnick & Bornstein, 2016), which are detailed in the following section. There was also an issue with the modification indices (MI), which are also based on the χ^2 statistic (Whittaker, 2012). Due to the sample-related inflation of the statistic, trivial changes were signaled as highly significant by the MIs, thus rendering the usual MI thresholds (Marôco, 2010) functionally useless. As a result, MIs were used in a limited manner. More details on how they were implemented are provided in the relevant section. Finally, the measurement invariance could not be tested using χ^2 comparisons, for the same reasons. Instead, AFIs were used (Meade & Lautenschlager, 2004; Putnick & Bornstein, 2016) in

accordance with the stated guidelines for best practice in the literature (Cheung & Rensvold, 2002; Milfont & Fischer, 2010).

3.3. Fit evaluation

Following model estimation, it is necessary to evaluate the model fit. Due to the large number of fit indicators, each representing different features of goodness-of-fit, it is usual to select one indicator for each category of indicators rather than report the entire suite of indicators (Bentler, 1990). The most common measure of fit is the χ^2 goodness-of-fit test (Barrett, 2007), which tests the null hypothesis that the population's covariance matrix is identical to the covariance matrix estimated by the model. However, due to the sample-related issues noted above, our evaluation relied heavily on the AFIs listed below.

The first category of fit indices is the *absolute indices*, which provide a measure of fit (Marôco, 2010). Traditionally, this is done using χ^2/df , the ratio of the chi-square statistic to the degrees of freedom. However, due to the large sample, it became necessary to use an alternative indicator for this category. We used the goodness-of-fit index (GFI), which is also commonly used in the literature. The second category of indices is the *comparative indices*, which compare the model fit with the fit of the independence and the saturated model (Bentler, 1990; Marôco, 2010). In this case, we used the comparative fit index (CFI) (Bentler, 1990). For the category of *parsimony-adjusted indices*, which penalize more complex models (Marôco, 2010), we used the parsimony-adjusted counterpart to the CFI, the PCFI. The fourth category comprised the *population-discrepancy indices*, which compare the model fit as calculated by the sample moments, where the model fit is calculated through population moments (Marôco, 2010). For this category we used the commonly used root mean square error of approximation (RMSEA), which is a popular choice because it is relatively insensitive to index inflation (Steiger, Shapiro, & Browne, 1985). The final category of *information-theory indices* is also dependent on the χ^2 statistic, but in this scenario this is less problematic as the values of these indices are devoid of meaning on their own. Rather, they are used to compare multiple models and are read as “less is better” (Anderson, Burnham, & White, 1998; Marôco, 2010). For this category, we used the modified expected cross-validation index (ECVI), which does not require the competing models to be nested (O’Rourke & Hatcher, 2013) and is considered to be particularly useful for CFA purposes (Bandalos, 1993). We used the modified version of ECVI because it is preferable under ML estimation (Marôco, 2010).

3.4. Modification Indices

To increment the model fit, it is possible to carry out model re-specifications. The first approach to re-specification eliminates non-significant trajectories and trajectories with low loadings, which has the additional advantage of increasing the factorial validity (Marôco, 2010). The second strategy involves MIs, which estimate the discrepancy or delta in the χ^2 statistic when certain adjustments are made to the model. It is important that these adjustments are coherent at a conceptual level, as otherwise a model can statistically have a good fit but be theoretically implausible (Arbuckle, 2007). This is usually performed by drawing covariances between error terms within the same factors, and eliminating variables with cross-loadings, which tend to manifest as high MI values connected to the covariances between error terms of variables in different factors (Marôco, 2010). In AMOS 24, the MIs use the Lagrange multipliers method (Bollen, 2014). MI analysis is usually conducted iteratively. The adjustments are first specified with an MI of 11 or higher, corresponding to a type I error probability of 0.001, and then with an MI of 4 or higher, representing a type I error probability of 0.05 (Marôco, 2010). In the main study, MIs were used sparsely due to the sample size.

3.5. Imputation

Missing values were imputed via Markov Chain Monte Carlo multiple imputation, which produced five complete datasets. EFA was carried out for each of the five complete datasets, and pooled estimates were then produced. In the CFA stage, because AMOS does not have built-in integration with the SPSS multiple imputation module, we used a single complete dataset.

3.6. Scale level

The original MDRAI and the new MDRAI-R items are scored on a 7-point Likert scale ranging from “completely disagree” to “completely agree.” Although Likert scales are technically ordinal, the data are treated as continuous throughout the entirety of the analysis. The rationale for this is as follows. First, various studies indicate that at the 5-point range and beyond, Likert scales can simply be treated as continuous (e.g., Johnson & Creech, 1983; Norman, 2010; Sullivan & Artino Jr, 2013; Zumbo & Zimmerman, 1993). In the context of SEM specifically, Kline (2016) only recommends using alternative estimation methods (i.e., not ML) when the range of the scale is 5 points or smaller. Indeed, this is precisely why we opted to use a 7-point scale, which is less common than the 5-point scale. Second, the skewness and kurtosis values for the individual items indicate that they are sufficient approximations of a normal distribution (as we demonstrate in a later section), further indicating that the items can reasonably be treated as continuous.

3.7. Procedures

We conducted several searches on the Scopus database from June 2017 to August 2018 to identify the corresponding authors of articles from all fields of knowledge (based on the Scopus disciplinary area classifications) published from 2010 to 2016. As the Scopus database only shows the results for the first 2,000 matches, several sorting strategies were used to maximize coverage, namely, default sorting, most relevant, least relevant, and highest cited. No further sorting strategies were used, as significant numbers of duplicate records had been obtained by this point. We found 915,447 corresponding authors.

The survey was carried out electronically through an online surveying platform. Invitations to participate were sent out by e-mail in batches from June 2017 to August 2018 (this included an additional wave of invitations to the authors that did not respond to the initial invitation). The invitation included a description of the project and the survey aims, and an opt-out link for participants who did not wish to be contacted again about the project. Those who accepted the invitation were directed to a page with an informed consent letter describing the scope, objectives, and purposes of the survey in further detail. The participants were required to give informed consent before they could proceed to the survey itself.

In total, 21,016 individuals agreed to participate. Of these, 8,883 dropped out before completion and were thus removed from the subsequent analysis. The final sample comprised 12,183 participants, of whom 4,153 (34.1%) were female and 8,030 (65.9%) were male. The mean age was 49.994 years ($SD = 12.285$). In regard to geographical distribution, the most represented countries were the United States ($N = 2235$; 18.3%), Italy ($N = 806$; 6.6%), the United Kingdom ($N = 760$; 6.2%), Spain ($N = 554$; 4.5%), and France ($N = 548$; 4.5%). The remaining participants were distributed across a range of other countries, ensuring global coverage. Table 1 summarizes the descriptive statistics for the sample. The geographical distribution is shown in Appendix A, due to its size.

Table 1: Descriptive statistics for the sample

Qualitative variables		N	%
Gender	Female	4153	34.1
	Male	8030	65.9
Field of knowledge	Natural and agricultural sciences	3309	27.2
	Engineering and technology	2553	21.0
	Medical and health sciences	3118	25.6
	Social sciences	2854	23.4
	Humanities	349	2.9
Quantitative variables		M	SD

Age	49.994	12.285
-----	--------	--------

Finally, for cross-validation purposes, the working dataset was randomly divided into two subsamples (see, e.g., Johnson & Stevens, 2001): a training dataset, comprising roughly 10% of the participants (N = 1203), to be used in the EFA, and a holdout dataset, comprising the remaining 90% of the participants (N = 10980), to be used for the CFA.

3.8. Analytical roadmap

We describe our analytical strategy as follows. We begin by reporting the results of a pilot study that was conducted prior to the main survey and the subsequent analysis. We then report the EFA results for the main study, which was conducted with the goal of obtaining a preliminary data structure for the new scales to be included in the model. EFA was followed by CFA, where the model was further refined through iterative re-specification until an optimal fit had been attained. After reporting the results of CFA, we describe the findings of our validity, reliability, and sensibility analyses, conducted to demonstrate the psychometric properties of the instrument. We conclude with measurement invariance analysis, which was performed to demonstrate that the instrument has similar measurement properties across all fields of knowledge.

4. Results

4.1. Pilot study

A pilot study was conducted in May and early June 2017 in preparation for the primary survey and the subsequent validation exercise. The pilot study aimed to (1) reinforce any weak pre-existing scales (i.e., those with the minimum number of items per dimension or items with relatively lower loadings in the MDRAI); (2) develop new questions related to entirely new themes that had emerged since the development of the original MDRAI; and (3) ensure that the global number of items was reasonable by filtering out unnecessary items without compromising the factorial structure (as an excessively lengthy survey can discourage participants from completing it).

A pool of 92 questions was developed based on these criteria. This pool comprised 22 items unchanged from the original MDRAI, and 13 items that were edited for clarity based on the comments by the participants in the pilot study. The 57 remaining items were original. Of these, 35 items were intended to reinforce the pre-existing scales, with the remaining 22 related to

novel themes, most notably orientation (towards institutions, community, or society) and external metric-driven pressure.

Participation in the pilot study was by invitation. The authors sent invitations to several researchers from a variety of fields of knowledge and institutions around the world. A public invitation was posted on the project's ResearchGate page. Ninety-seven researchers agreed to participate in the pilot study. The questions were presented in random order to each participant.

The data obtained in the pilot study were analyzed using EFA and then CFA. Each scale was analyzed independently due to a) the small sample size for the pilot study and b) the expectation of relative independence for each scale (they are meant to be able to be used individually if desired, as each scale measures a separate facet of a research agenda). For the new themes, EFA was conducted using Varimax rotation (Ebrahimi & Osareh, 2014), and the optimal number of factors was determined using the following criteria: a) Kaiser's criterion, b) the scree-plot's "elbow," and c) the percentage of extracted variance. The extracted structure was then specified in the CFA stage for further evaluation.

The two main conclusions of this study relate to the new themes. The item elimination, although necessary, was less interesting and the results are summarized in a later section. The items related to the new orientation scale originally revealed three factors explaining 67.38% of the variance. Based on their content, the items seemed to be related to the field orientation (e.g., "My choice of topic is determined by the field community"), society orientation (e.g., "I decide my research topic based on societal challenges"), and institutional orientation (e.g., "My research agenda is aligned with my institution's research strategies"). Thus, the CFA specified a model with three lower-order latent variables in accordance with this structure. The field and institutional orientation dimensions had reasonable loadings (0.72) and (0.91), but society orientation loaded poorly onto the higher order factor (0.39). We interpreted this as indicative that a society orientation can sometimes be at odds with an academic orientation, or in practical terms, that the society orientation factor might be independent of the other two orientation factors. We decided to reinforce the society orientation factor (which had only three items) with an additional three items and repeat the EFA for this factor in the main study. This generated two new sub-scales: one comprising the field and institutional orientation scales (which we termed "Academia Driven"), and a second comprising the society-related items (which we termed "Society Driven"). Our second conclusion concerns the metric-driven pressure scale, which identified two factors explaining 55.87% of the variance: one related to publication pressure and the other to evaluation metrics pressure. This sub-scale was tentatively termed

“Publish or Perish.” The pilot study concluded with a preliminary version of the revised survey comprising 68 items, which was used in the main study as described below.

4.2. EFA

Before conducting the CFA, a new EFA was conducted on the new scales (Academia Driven, Society Driven, and Publish or Perish) using the training dataset, similar to the EFA in the pilot study, to obtain a tentative factorial structure for the CFA stage (Bentler & Weeks, 1980). Accordingly, three independent EFAs were conducted, one for each scale. Although we could have conducted a single EFA, we decided to use identical procedures to the pilot to ensure consistency and reflect the modular nature of the inventory.

The EFA for the Academia Driven sub-scale largely matched that observed during the pilot study, with two extracted factors explaining 69.43% of the variance. Semantic interpretation of the items loading onto each factor exhibited similar behavior to that previously observed, with a factor related to institutional orientation and another to field orientation. The Society Driven scale, with the reinforcement items added in the previous stage, showed that two factors explained 79.26% of the variance. Semantic analysis of the items suggested that one of the factors was related to society (e.g., “I decide my research topics based on societal challenges”), while another was related to interactions with non-academics (e.g., “I choose my research topics based on my interaction with my non-academic peers”). We tentatively named these two factors “Society” and “Non-academic.” Finally, in contrast with the observations from the pilot study, the Publish or Perish scale revealed that a single item explained 47.77% of the variance. Due to the previous findings and because the analysis scree plot suggested a possible 2-factor solution, a forced 2-factor extraction was attempted. However, this revealed significant cross-loadings on both factors from multiple items, thus confirming that the 1-factor solution was optimal. As such, we decided to use the 1-factor solution in the CFA stage, and re-evaluate the structure of this scale based on the findings. Table 2 summarizes the results of these analyses.

Table 2: Exploratory Factor Analysis for the new scales

	Factor	
	1	2
Academia Driven scale		
My choice of topics is determined by my field community.	0.1524	0.7478
I adjust my research agenda based on my institution's demands.	0.8478	0.1976
My research agenda is aligned with my institution's research strategies.	0.7896	0.2072
My institution defines my research agenda.	0.8356	0.2118
I often decide my research agenda in collaboration with my field community.	0.2106	0.8044
My research agenda depends on the field community.	0.2388	0.8274
Society Driven scale		
I decide my research topic based on societal challenges.	0.8672	0.1960
I choose my research topics based on my interactions with my non-academic peers.	0.2434	0.8212
I consider my research topics myself, but this consideration often occurs after I hear what my non-academic peers have to say about these topics.	0.1378	0.8520
Societal challenges drive my research choices.	0.8854	0.2534
I often strive to engage in issues that address societal challenges.	0.8716	0.2278
I consider the opinions of my non-academic peers when I choose my research topics.	0.2876	0.8362
Publish or Perish scale		
	1	
I do not choose research topics that receive poor project evaluations.	0.6538	
I often choose research topics that lead to many publications.	0.6072	
Publish or perish defines my research agenda.	0.7016	
If research topics do not warrant the potential for many publications and citations, I do not choose them.	0.7488	
My choice of research topics is aligned with expected research evaluations.	0.7496	
My work is constrained by evaluation frameworks.	0.5924	
Note: standardized loadings from Varimax rotation are reported. Bolded values indicate the factor with the highest loading.		

4.3. Model Specification

From this section onwards, the holdout sample is used for the reported analysis. The initial specification strategy replicated the structure obtained during the CFA for the original version of the instrument for the changed scales (Horta & Santos, 2016), and replicated the structure obtained during the EFA stage (see the previous sub-section) for the new scales (Marôco, 2010). This specification resulted in a model with an inadmissible solution due to a non-positive definite covariance matrix. This is a difficult issue to address, as it does not have a clear cause or method of diagnosis. In the literature, this is attributed to small sample sizes, insufficient numbers of manifest variables for each latent variable, misspecification of the model, and multicollinearity (Hair et al., 2007; Kline, 2016; Marôco, 2010). However, the issue needed to be resolved before proceeding with the analysis. As the sample for this exercise was not small and the recommended number of items per latent variable was met or exceeded in each case (Marôco, 2010), the only plausible remaining solutions were misspecification of the model or

multicollinearity. As this was a CFA exercise, rather than path analysis, multicollinearity was somewhat expected and desirable (despite conceptual expectations of varying degrees of independence of some of the scales). Nevertheless, we speculated that there could be some degree of overlap leading to a misspecification issue. To diagnose this, we re-ran an EFA, but this time with the entire pool of items. The issue then became apparent. In the original validation exercise, some competing dimensions had loaded onto separate factors (Horta & Santos, 2016), but in this exercise they exhibited different behaviors. Some of the items in the Conservative scale loaded onto the same factor as the Convergence scale, while some items for the Convergence scale loaded onto the Divergence scale, albeit with a negative loading, while simultaneously exhibiting cross-loadings with the remaining items of the Convergence scale. This strongly suggests the redundancy of these scales, in the sense that Convergence/Divergence and Discovery/Conservative can be measured on a spectrum using a single scale rather than independent scales. As such, it was decided to remove the Convergence and Conservative scales entirely and instead measure these concepts through the Divergence and Discovery scales (i.e., lower scores for Divergence translate to higher scores for Convergence characteristics). An additional issue emerged in the new Publish or Perish scale, which exhibited substantial cross-loadings across the board, and thus was considered unviable for inclusion in the instrument. The removal of these scales addressed the issue and allowed an admissible solution to be estimated. An incidental benefit was that this further assisted the stated goal of reducing the number of items in the instrument.

The second step for specification was scanning for items with poor loadings (under 0.50), which indicate poor factorial validity (Kline, 2016; Marôco, 2010). The only such item was one of the new items in the Discovery scale (“I invest most of my time in research that I believe is at the forefront of knowledge”), with $\lambda = 0.44$. All of the other items were above the required threshold. This item was removed, and the model was re-estimated.

The third step involved removing redundant items, in line with the stated goal of reducing the number of items. The main candidate scales for item reduction were Mentor Influence, Tolerance of Low Funding, and Discovery, all with six items each. Observing the MIs, it was evident that there were substantial within-scale correlations between the error terms for the respective items, suggesting the redundancy of some of these items and providing grounds for their removal. Although there is no consensus on the optimal number of items for measuring a latent factor, similar analyses have been carried out with as few as two manifest variables (Rammstedt & John, 2007). However, most scholars consider this to be the absolute minimum, with a recommended minimum of three (Hair et al., 2007; Marôco, 2010). We opted to reduce

Copyright: © 2019 Hugo Horta and João M. Santos. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

the number of items in these scales to four. We decided to remove the two worst performing items in each of the scales (due to either poor loadings or high cross-loadings). For the Tolerance of Low Funding scale, the two items removed were “I try not to worry about funding availability when I plan my research,” with $\lambda = 0.65$, and “I think I can progress in my career doing research with limited funding,” with $\lambda = 0.58$. For the Discovery scale, the items were “I have a preference for new research topics,” with $\lambda = 0.62$, and “I prefer to work on topics that have a high degree of novelty,” with $\lambda = 0.77$. Finally, for Mentor Influence, the removed items were “My PhD mentor’s opinion carries much weight in my research choices,” with $\lambda = 0.71$, and “My PhD mentor still often works alongside me,” with $\lambda = 0.69$. In addition, one of the items on the Prestige sub-scale of the Scientific Ambition scale (“Standing out from the rest of my peers is one of my goals”) performed somewhat worse than its peers, with $\lambda = 0.68$. As the Scientific Ambition scale was already measured by seven items (four for Prestige and three for Drive to Publish), we decided to also remove this item. After this round of removals, the model was re-estimated.

The fourth and final step was evaluating the MIs. This was a daunting task, as MI values are based on the χ^2 statistic (Whittaker, 2012). As noted in the methods section, this statistic was substantially inflated by the sample size, which also caused the MIs to be inflated by proxy, resulting in trivial model changes to be flagged as highly significant by the MIs. Specifically, the threshold value of 11, which corresponds to a Type I error probability of 0.001 (Marôco, 2010), applied to nearly all of the proposed changes. We opted to implement modifications following the usual convention of creating covariances between error terms loading onto the same factor (Kline, 2016; Marôco, 2010), and evaluate the effective fit gain through the AFIs. Other than the within-factor error disturbances, two items were removed due to substantial cross-loadings evident from very high MI values, both from the Academia Driven scale: “I often decide my research agenda in collaboration with my field community” and “My institution defines my research agenda.” As the χ^2 statistic could not be used to gauge the quality of the model changes, we opted to evaluate improvements through the CFI instead. For each implemented MI change, the model was re-estimated and re-evaluated in an iterative manner until a CFI above 0.950 was reached. This level is considered the highest possible qualitative threshold for model fit using this index (Hu & Bentler, 1999).

This multi-stage specification strategy yielded notable gains in model fit ($MECVI_{\text{initial}} = 1.941$ versus $MECVI_{\text{final}} = 1.103$), accomplished the goal of item reduction, and addressed all of the specification issues. The fit evaluation at each stage is summarized in Table 3.

Table 3: Model fit evaluation

<i>Model</i>	<i>GFI</i>	<i>CFI</i>	<i>PCFI</i>	<i>RMSEA</i>	<i>MECVI</i>
I	0.921	0.929	0.862	0.040	1.941
II	0.924	0.932	0.863	0.040	1.841
III	0.938	0.942	0.855	0.040	1.315
IV	0.950	0.953	0.850	0.037	1.103

Notes: Model I: initial admissible model; Model II: model without items with poor loading; Model III: model without redundant items; Model IV: model with MI implementations.

4.4. CFA

Full information ML was used to estimate the final model. For this final iteration, the model was as significant as the various trajectories ($p < 0.001$). Based on the fit evaluation, and using the common thresholds (Barrett, 2007; Hair et al., 2007; Hooper, Coughlan, & Mullen, 2008; Kline, 2016; Marôco, 2010), it was determined that the model fit could be qualitatively assessed as very good ($GFI = 0.950$; $CFI = 0.953$; $PCFI = 0.850$; $RMSEA = 0.037$). Table 4 indicates the factorial loadings for the final model, and Figure 1 provides a visual representation of the model. Finally, Table 5 provides item-level descriptive statistics; from which it can also be observed that all of the items follow univariate normality, following Kline's (2016) criteria for skewness and the kurtosis thresholds.

Table 4: Factorial loadings for the MDRAI-R

Code	Item	Loading
A1	I aim to one day be one of the most respected experts in my field.	0.802
A2	Being a highly regarded expert is one of my career goals.	0.802
A3	I aim to be recognized by my peers.	0.704
A5	I feel the need to constantly publish new and interesting papers.	0.782
A6	I am constantly striving to publish new papers.	0.873
A7	I am driven to publish papers.	0.792
DV1	I look forward to diversifying into other fields.	0.720
DV2	I would be interested in pursuing research in other fields.	0.781
DV4	I would like to publish in different fields.	0.737
DV5	I enjoy multi-disciplinary research more than single-disciplinary research.	0.851
DV6	Multi-disciplinary research is more interesting than single-disciplinary research.	0.877
DV8	I prefer to work with multi-disciplinary rather than single-disciplinary teams.	0.874
COL2	My publications are enhanced by collaboration with other authors.	0.604
COL4	I often seek peers with whom I can collaborate on publications.	0.655
COL5	I enjoy conducting collaborative research with my peers.	0.734
COL7	My peers often seek to collaborate with me in their publications.	0.741
COL8	I am often invited to collaborate with my peers.	0.908
COL12	I am frequently invited to participate in research collaborations due to my reputation.	0.827
M2	Part of my work is largely due to my PhD mentor.	0.787
M3	My research choices are highly influenced by my PhD mentor's opinion.	0.852
M4	My PhD mentor is responsible for a large part of my work.	0.892
M6	My PhD mentor largely determines my research topics.	0.931
TTLF1	Limited funding does not constrain my choice of topic.	0.822
TTLF2	Highly limited funding does not constrain my choice of topic.	0.865
TTLF3	The availability of research funding for a certain topic does not influence my decision to conduct research on that topic.	0.696
TTLF4	I am not discouraged by the lack of funding on a certain topic.	0.616
D2	I would rather conduct revolutionary research with little chance of success than replicate research with a high probability of success.	0.684
D3	I prefer "innovative" research to "safe" research, even when the odds of success are much lower.	0.687
D4	I would rather engage in new research endeavors, even when success is unlikely, than safe research that contributes little to the field.	0.701
D9	I am driven by innovative research.	0.678
O1	My choice of topics is determined by my field community.	0.600
O9	I often decide my research agenda in collaboration with my field community.	0.803
O6	I adjust my research agenda based on my institution's demands.	0.759
O7	My research agenda is aligned with my institution's research strategies.	0.733
S1	I decide my research topic based on societal challenges.	0.807
S4	Societal challenges drive my research choices.	0.904
S5	I often strive to engage in issues that address societal challenges.	0.843
S2	I choose my research topics based on my interactions with my non-academic peers.	0.769
S3	I consider my research topics myself, but this consideration often occurs after I hear what my non-academic peers have to say about these topics.	0.732
S6	I consider the opinions of my non-academic peers when I choose my research topics.	0.868

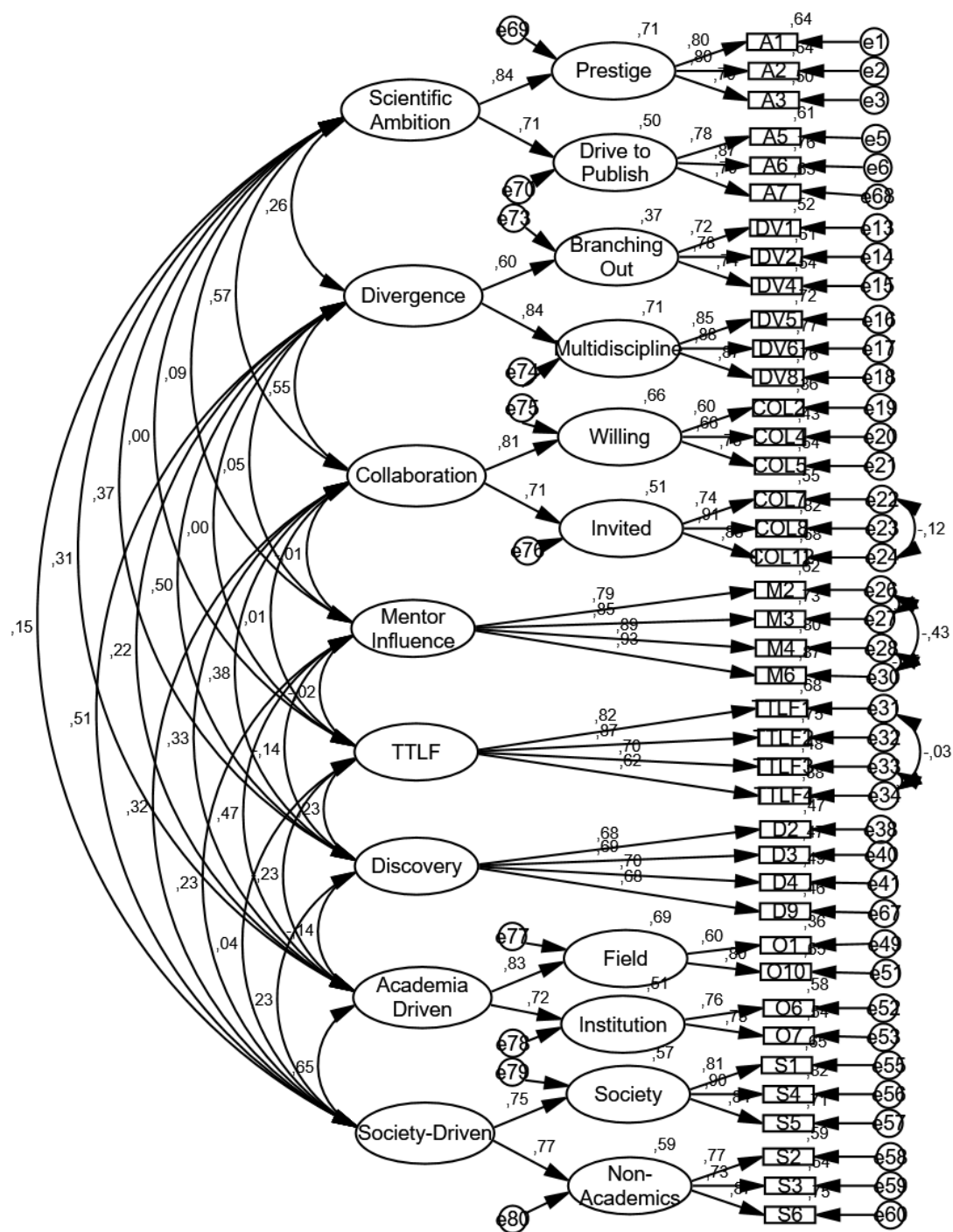


Fig. 1. CFA model for the MDRAI-R.

Table 5: Item-level descriptive statistics

Code	Item	Mean	Std Dev	Skewness	Kurtosis
A1	I aim to one day be one of the most respected (...)	4.858	1.380	-0.364	-0.003
A2	Being a highly regarded expert is one of my (...)	5.168	1.321	-0.615	0.343
A3	I aim to be recognized by my peers.	5.135	1.142	-0.559	0.993
A5	I feel the need to constantly publish new (...)	5.046	1.243	-0.553	0.448
A6	I am constantly striving to publish new papers.	5.009	1.237	-0.473	0.373
A7	I am driven to publish papers.	5.033	1.265	-0.595	0.621
DV1	I look forward to diversifying into other fields.	4.987	1.286	-0.515	0.272
DV2	I would be interested in pursuing research (...)	4.818	1.205	-0.498	0.553
DV4	I would like to publish in different fields.	4.769	1.166	-0.431	0.546
DV5	I enjoy multi-disciplinary research more than (...)	5.305	1.227	-0.498	0.151
DV6	Multi-disciplinary research is more interesting (...)	5.297	1.209	-0.513	0.364
DV8	I prefer to work with multi-disciplinary rather (...)	5.136	1.238	-0.431	0.194
COL2	My publications are enhanced by collaboration (...)	5.885	1.179	-1.257	2.062
COL4	I often seek peers with whom I can collaborate (...)	4.919	1.275	-0.484	0.244
COL5	I enjoy conducting collaborative research with (...)	5.572	0.990	-0.503	1.093
COL7	My peers often seek to collaborate with me in (...)	4.832	1.143	-0.391	0.535
COL8	I am often invited to collaborate with my peers.	4.901	1.126	-0.445	0.734
COL12	I am frequently invited to participate in (...)	4.777	1.200	-0.394	0.409
M2	Part of my work is largely due to my PhD mentor.	3.403	1.678	0.166	-0.829
M3	My research choices are highly influenced by (...)	3.063	1.567	0.355	-0.537
M4	My PhD mentor is responsible for a large part (...)	2.776	1.543	0.554	-0.323
M6	My PhD mentor largely determines my (...)	2.645	1.504	0.607	-0.205
TTLF1	Limited funding does not constrain my choice (...)	4.119	1.761	-0.016	-0.986
TTLF2	Highly limited funding does not constrain my (...)	4.033	1.660	-0.005	-0.856
TTLF3	The availability of research funding for (...)	4.214	1.499	-0.022	-0.644
TTLF4	I am not discouraged by the lack of funding on (...)	4.475	1.449	-0.327	-0.356
D2	I would rather conduct revolutionary research (...)	4.812	1.273	-0.112	-0.271
D3	I prefer “innovative” research to “safe” (...)	5.101	1.223	-0.379	0.008
D4	I would rather engage in new research (...)	4.937	1.167	-0.249	0.158
D9	I am driven by innovative research.	5.237	1.025	-0.260	0.542
O1	My choice of topics is determined by my field (...)	4.220	1.414	-0.229	-0.368
O9	I often decide my research agenda in (...)	4.271	1.271	-0.384	0.050
O6	I adjust my research agenda based on my (...)	3.789	1.525	-0.062	-0.696
O7	My research agenda is aligned with my (...)	4.253	1.393	-0.364	-0.135
S1	I decide my research topic based on societal (...)	4.545	1.509	-0.431	-0.297
S4	Societal challenges drive my research choices.	4.452	1.453	-0.454	-0.173
S5	I often strive to engage in issues that address (...)	4.564	1.404	-0.442	-0.033
S2	I choose my research topics based on my (...)	3.623	1.422	0.030	-0.444
S3	I consider my research topics myself, but (...)	3.644	1.412	-0.026	-0.483
S6	I consider the opinions of my non-academic (...)	3.754	1.426	-0.162	-0.502

552

553 In addition to the factorial loadings, initial insights regarding the interplay of the various
554 dimensions can be obtained by observing the correlations in Figure 1. First, a moderately strong
555 correlation can be observed between the Academia-Driven and Society-Driven scales ($r = 0.646$).
556 A possible explanation is that institutions (and indeed the academy) currently place emphasis
557 on society-focused research, causing them to be somewhat aligned, even if they are still
558 independent (and, as mentioned in the pilot study section, sometimes at odds with each other).
559 The Society-Driven scale also exhibits a moderate correlation with Divergence ($r = 0.508$), which

suggests either that the society-focused challenges are requiring more multidisciplinary approaches or that researchers who have a preference for diverging research are also more likely to engage in society-driven research. Divergence also exhibits a moderate correlation with Discovery ($r = 0.503$), which is expected because these two agendas are core traits of the trailblazing doctrine that was identified in the previous iteration of the MDRAI (Santos & Horta, 2018). Similarly, Collaboration exhibits moderate correlations with Scientific Ambition ($r = 0.568$) and Divergence ($r = 0.554$), and thus also resonates with the characteristics of the trailblazing doctrine. Several other correlations, which are not covered here but are relatively easy to interpret, can be identified, but they are not as strong. Overall, the observed correlational matrix can provide insights into how to use the MDRAI-R in future studies.

4.5. Validity, Reliability, and Sensitivity

Three types of validity were assessed in this study: factorial validity, convergent validity, and discriminant validity (Hair et al., 2007; Marôco, 2010). James Gaskin's Stats Tool Package (2016), specifically the Validity Master macro, was used for the assessment. This also reflects the same types of validity evaluated in the validation exercise for the first version of the MDRAI. *Factorial validity* can be attained when the standardized loadings for all items exceed the 0.50 threshold (Marôco, 2010). One of the steps in the previous section ensured that this criterion was met, so the model had factorial validity.

The second type, *convergent validity*, relates to high loadings from the manifest variables onto the latent variables, and is evaluated through the average variance extracted (Fornell & Larcker, 1981). The AVE for a given factor is given by:

$$\widehat{AVE}_j = \frac{\sum_{i=1}^k \lambda_{ij}^2}{\sum_{i=1}^k \lambda_{ij}^2 + \sum_{i=1}^k \varepsilon_{ij}}$$

Based on this calculation, convergent validity is confirmed when the AVE exceeds the 0.50 threshold (Hair et al., 2007). This was the case for all of the factors, with the exception of Discovery, with a slightly lower AVE of 0.473. Although this could conceivably have been increased by eliminating the lowest-loading item, a minor shift from the threshold is likely to be irrelevant at a practical level. Therefore, we argue that convergent validity was largely demonstrated, although the abovementioned issue must be taken into consideration when using the Discovery scale. We proceeded by evaluating the *discriminant validity*, which reflects the degree of extra-factorial correlation. Discriminant validity is demonstrated when the square root of the AVE for a given pair of factors i and j is equal to or greater than the correlations

between those two factors. Furthermore, the AVE must be equal to or greater than both the maximum shared variance (MSV) and the average shared variance (ASV) (Fornell & Larcker, 1981; Hair et al., 2007; Marôco, 2010). All of the factors met this criterion, demonstrating the discriminant validity of the instrument.

Following the validity evaluation, we proceeded with the analysis of *reliability*, which is a measure of consistency (Marôco, 2010). This was done using the composite reliability (CR) (Fornell & Larcker, 1981), which for a given factor j with k items is given by:

$$\widehat{CR}_j = \frac{(\sum_{i=1}^k \lambda_{ij})^2}{(\sum_{i=1}^k \lambda_{ij})^2 + \sum_{i=1}^k \varepsilon_{ij}}$$

The proposed threshold of 0.7 is considered to indicate scale reliability (Hair et al., 2007). All of the factors exceeded the required threshold, with the exception of Divergence (CR = 0.695). However, as before, a millesimal difference is likely to be trivial. Despite this slight deviation, the instrument can be considered reliable. Table 6 summarizes the validity and reliability findings, and the correlations between the factors.

Table 6: Validity and Reliability evaluation

	CR	AVE	MSV	ASV	Correlations							
					Ambi	Acad	Soci	Disc	TTLF	Ment	Coll	Dive
Ambition	0.751	0.604	0.323	0.094	0.777							
Academia	0.750	0.601	0.417	0.138	0.314	0.775						
Society	0.732	0.577	0.417	0.130	0.152	0.646	0.760					
Discovery	0.782	0.473	0.253	0.097	0.370	-0.139	0.234	0.688				
TTLF	0.840	0.572	0.052	0.015	-0.004	-0.229	0.037	0.228	0.756			
Mentor	0.924	0.752	0.222	0.043	0.094	0.471	0.227	-0.136	-0.024	0.867		
Collab	0.738	0.586	0.323	0.141	0.568	0.331	0.323	0.381	0.013	-0.010	0.766	
Divergence	0.695	0.539	0.307	0.134	0.257	0.219	0.508	0.503	-0.003	0.052	0.554	0.734

Note: the diagonal of the correlation matrix indicates the square root of the AVE.

The final factor is *sensitivity*, which refers to the capability of an instrument to differentiate between two individual items. This is demonstrated when all of the individual items have a reasonably normal distribution (Marôco, 2010). Items are considered to have a reasonable approximation to the normal distribution when their skewness and kurtosis are under the absolute value of 3 (Kline, 2016). All of the items were below this threshold for both parameters, thus demonstrating the sensitivity of the instrument and completing the validation exercise.

4.6. Measurement Invariance

In this step, the goal was to assess and eventually demonstrate measurement invariance across the major fields of knowledge. The fields of knowledge were the Exact and Natural Sciences, Health and Medical Sciences, Engineering and Technology, Social Sciences, and Humanities. Measurement invariance indicates that the operationalization of a construct has the same meaning in different contexts (Meade & Lautenschlager, 2004). In other words, its metric is universal wherever invariance is tested. To achieve this, we used a multi-group analysis following the procedure outlined by Marôco (2010) and Kline (2011), which involves comparing the unconstrained model with progressively more constrained models. Typically, this is done using χ^2 tests for difference. However, as noted in the literature and observed in our own dataset, this statistic becomes unreliable with larger samples, as all trivial differences are deemed to be significant (Chen, 2007; Cheung & Rensvold, 2002; Kline, 2016; Meade, Johnson, & Braddy, 2008; Putnick & Bornstein, 2016). Scholars have proposed using AFI in these scenarios instead (Putnick & Bornstein, 2016). Cheung and Rensvold (2002) propose that a CFI change of less than 0.01 indicates measurement invariance. Thus, we estimated the multi-group analysis for fields of knowledge using progressive levels of constraints, based on the hypotheses for testing measurement invariance proposed by Cheung and Rensvold (2002) and following the guidelines recommended by Milfont and Fischer (2010).

We began by testing hypothesis H_{λ} . Metric invariance was demonstrated for the first model, with a ΔCFI of 0.000 (Model II), indicating that the constructs manifest identically across fields of knowledge (Cheung & Rensvold, 2002). For the next hypothesis, $H_{\lambda, \theta(\delta)}$, residual variances and covariances were also demonstrated, with a ΔCFI of 0.002 (Model III), indicating that the internal consistency is identical across the fields of knowledge (Cheung & Rensvold, 2002). The threshold for hypothesis $H_{\lambda, \nu}$, scalar invariance, was not met, with a ΔCFI of 0.012 (Model IV). Following the guidelines in the literature for best practice in testing measurement invariance, we then tested for partial scalar invariance (Byrne, Shavelson, & Muthén, 1989; Cheung & Rensvold, 2002; Milfont & Fischer, 2010). This required us to determine which intercepts varied to the greatest degree across the fields of knowledge. Due to the large number of intercepts and groups, a more efficient method than simple visual inspection of the intercept matrix was required. We computed the square root of the sum of the squared differences for each pair of intercepts to identify which intercepts had the largest cross-field of knowledge discrepancies. These intercepts lay in two scales: Tolerance of Low Funding and the new Society Driven scale. This finding can be explained as follows. For Tolerance of Low Funding, it could relate to the widely varying availability of funding across the fields of knowledge, leading to different levels of risk tolerance (Lanahan, Graddy-Reed, & Feldman, 2016; Mejia & Kajikawa, 2018). Similarly,

Copyright: © 2019 Hugo Horta and João M. Santos. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

for the Society Driven scale, the finding could relate to the difference between basic and applied research, as basic research has lower levels of Society Driven research agenda characteristics than applied research (see Bentley, Gulbrandsen, & Kyvik, 2015).

Having identified the source of variance, we allowed these intercepts to vary freely across the fields of knowledge and proceeded with the analysis, as per the guidelines provided by Milfont and Fischer (2010). The new model met the threshold for partial scalar invariance, with a ΔCFI of 0.008 (Model V), indicating scalar invariance for all of the scales except Tolerance of Low Funding and Society Driven. The next level of invariance is at the construct level. The next model constrained the construct variances and covariances, and tested hypotheses $H_{\Lambda, \theta(jj)}$ and $H_{\Lambda, \theta(jj')}$. The equivalence of construct variance and covariance was not demonstrated, with a ΔCFI of 0.011 (Model VI), indicating that the range of responses and relationships between the constructs were not identical across the groups. Finally, the last model tested for differences in the latent means (hypothesis $H_{\Lambda, \nu, \kappa}$), which were demonstrated with a ΔCFI of 0.000 (Model VII). As such, measurement invariance was demonstrated for the instrument, with full metric, scalar, and partial construct invariance for all of the scales except Tolerance of Low Funding and Society Driven, which nevertheless still possessed metric invariance. The results of the model comparison are summarized in Table 7. Finally, the descriptive statistics for each factor and for each field of knowledge are presented in Table 8.

Table 7: Comparison of unconstrained and constrained models

<i>Model</i>	<i>Hypothesis</i>	<i>Constraints</i>	<i>Level</i>	<i>CFI</i>	<i>ΔCFI</i>
I		Unconstrained	-	0.946	-
II	H_{λ}	Factor loadings	Metric	0.946	0.000
III	$H_{\Lambda, \theta(\delta)}$	Residuals (co)variance	Metric	0.944	0.002
IV	$H_{\Lambda, \nu}$	Intercepts	Scalar	0.932	0.012
V	$H_{\Lambda, \nu}$	Intercepts (Partial)	Scalar	0.938	0.008
VI	$H_{\Lambda, \theta(jj)} / H_{\Lambda, \theta(jj')}$	Construct (co)variance	Construct	0.935	0.011
VII	$H_{\Lambda, \nu, \kappa}$	Latent means	Construct	0.935	0.003

Notes: ΔCFI is calculated with reference to less constrained models using the guidelines in Cheung and Rensvold (2002). Although Cheung and Rensvold (2002) indicate equivalence of construct variance and equivalence of construct covariance as separate hypotheses, they were merged for this exercise. This is a technical limitation as the AMOS software package bundles these two constraints together.

Table 8: Descriptive statistics for all factors and sub-factors across fields of science

Factor	N&A		E&T		M&H		SS		H	
	M	SD	M	SD	M	SD	M	SD	M	SD
Scientific Ambition	5.017	0.945	5.007	0.953	5.068	0.916	5.071	0.993	5.054	0.956
Prestige	5.001	1.107	5.084	1.078	5.103	1.053	5.029	1.126	5.100	1.139
Drive to Publish	5.033	1.085	4.929	1.110	5.033	1.066	5.112	1.129	5.008	1.120
Divergence	5.002	0.934	5.195	0.858	5.060	0.849	4.971	0.959	5.054	0.958
Branching Out	4.858	1.010	5.025	0.959	4.730	1.026	4.834	1.062	4.953	1.046
Multidisciplinarity	5.147	1.163	5.365	1.050	5.391	1.043	5.108	1.169	5.155	1.167
Collaboration	5.133	0.824	5.040	0.768	5.289	0.779	5.144	0.868	4.861	1.043
Willing to Collab.	5.432	0.913	5.369	0.851	5.593	0.848	5.470	0.964	5.083	1.173
Invited to Collab.	4.833	1.012	4.711	0.963	4.985	0.960	4.818	1.091	4.638	1.167
Mentor Influence	2.847	1.373	3.134	1.420	3.085	1.384	2.879	1.439	2.701	1.411
TTLF	4.138	1.302	4.185	1.255	3.975	1.278	4.501	1.361	4.752	1.343
Discovery	5.010	0.927	5.080	0.864	4.971	0.894	5.026	0.952	5.113	0.934
Academia Driven	3.970	1.052	4.193	0.966	4.226	0.970	3.892	1.039	3.702	1.117
Field	3.982	1.163	4.199	1.088	4.269	1.116	3.963	1.171	3.819	1.270
Institution	3.959	1.326	4.186	1.234	4.184	1.224	3.822	1.304	3.585	1.349
Society Driven	3.658	1.187	4.204	0.975	4.208	0.989	4.373	1.063	4.170	1.125
Society	4.021	1.421	4.532	1.192	4.639	1.197	4.939	1.221	4.631	1.389
Non-Academics	3.296	1.244	3.876	1.149	3.777	1.165	3.807	1.260	3.709	1.206

Notes: N&A – Natural and Agricultural Sciences; E&T – Engineering and Technology; M&H – Medical and Health sciences; SS – Social Sciences; H – Humanities.

5. Discussion

In this section, the various scales and their scoring are interpreted. We first discuss the scales, and then focus on the scoring. To calculate composite scores for each scale, there are numerous options to choose from (DiStefano, Zhu, & Mindrila, 2009). As with the initial version of the MDRAI, simple summation is discouraged due to the unbalanced number of items across the factors. Although this was one of the goals of this revision, it was unfortunately not possible to do so and maintain the validity of the scale. Therefore, the score range varied across the scales, making direct comparison difficult. The simplest alternative way of computing the composite scores, and the approach we encourage for general use, is to calculate the mean score of the items in each scale. This yields a composite non-discrete score ranging from 1 to 7. In addition, the mean for each item can be weighted using the factor loadings provided in Table 4. Scores can be computed either for the first-order factors or for the second-order factors, depending on the specific research purposes.

The Scientific Ambition dimension retained the same importance as it had in the MDRAI, including that of its sub-dimensions (Prestige and Drive to Publish), stressing the relevance of

engaging in research agendas that can provide recognition for one's work from peers and help to achieve positions of intellectual and field authority in the knowledge communities of interest (Latour & Woolgar, 2013; Whitley, 2000). The Collaboration dimension and its sub-dimensions (Willing to Collaborate and Invited to Collaborate) also retained critical importance in the MDRAI-R, which demonstrates an understanding that collaborative agendas are necessary in all fields of knowledge, and that collaborating or not with peers is a key decision when embarking on new research agendas (Siciliano, Welch, & Feeney, 2018). Higher scores for the dimensions and the respective sub-dimensions mean that the relevance of these factors to the research agenda is more important for researchers, e.g. a higher score for Scientific Ambition means that researchers privilege this dimension when developing their research agendas.

The Tolerance of Low Funding and Mentor Influence dimensions also appear to be critical in influencing the research agendas of researchers, as they were in the MDRAI. Higher scores for Tolerance of Low Funding indicate that researchers are not discouraged by a lack of available funding from pursuing specific research agendas, meaning that they do not place an emphasis on research funding when deciding on a research agenda, while lower scores for this dimension indicate that researchers consider research funding to be a critical element when deciding on specific research agendas. We further argue that a median score in this dimension can indicate that in some cases researchers follow research funding when opting for specific research agendas, but not in others. This scoring could also indicate that researchers are willing to engage in exploratory research agendas that have little to no funding as a way to obtain initial findings that could allow them to then prepare research agendas of greater scope, ambition, and focus that might need research funding to come to fruition. A higher score for Mentor's Influence suggests that the PhD supervisor continues to have a say in or a degree of influence on a researcher's research agenda, while the opposite means that the researcher embarks on research agendas without requesting their PhD supervisor's guidance or opinion. These scores can be a proxy for researcher independence, but can also be understood as a measure of a researcher's relationship with his or her PhD supervisor after completing a doctorate (Ooms et al., 2018).

The Discovery dimension in the MDRAI-R combines the MDRAI dimensions of Discovery and Conservative into a single dimension, as discussed in the main study section, thereby placing the previously independent dimensions on a continuum. The higher the score for the Discovery dimension, the more likely the researcher is to engage in research agendas that are riskier, and focus on emerging and unexplored themes that have greater potential for breakthroughs but also for failure. Santos and Horta (2018) characterize researchers with high Discovery score

Copyright: © 2019 Hugo Horta and João M. Santos. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

research agendas as trailblazers, and Foster et al. (2015) characterize these researchers as having innovative research strategies. A lower score in this dimension indicates a preference for low risk research agendas that are more focused on the gradual accumulation of knowledge in well-established themes, topics, and fields. Santos and Horta (2018) characterize researchers with low Discovery score research agendas as cohesive, and Foster et al. (2015) characterize them as having traditional research strategies. The Divergence dimension maintained the same structure as in the MDRAI, including its sub-dimensions (Branching out and Multidisciplinary), but similar to the Discovery dimension, it also combined the MDRAI dimensions of Divergence and Convergence into a one-dimension continuum. A higher score in the Divergence dimension means that researchers establish research agendas that link and involve knowledge from other fields of knowledge, and are attuned to the current needs of complex problems (Zuo & Zhao, 2018). Lower scores in this dimension indicate research agendas bounded by a single field of knowledge, and are associated with specialization, knowledge mastery, field identity, and a focus on one or few topics rather than diversification (Franzoni & Rossi-Lamastra, 2017).

The first new dimension of the MDRAI-R is Academia Driven, which refers to the extent to which a research agenda is influenced by holistic, valuative, and normative traits and dispositions related to the scholarly and academic environment and social structure with which the researcher identifies. The higher the score in this dimension, the more the research agenda conforms to and is aligned with the questions, topics, and strategic focuses that the academic environment might regard as a priority. A lower score in this dimension indicates that a research agenda is more based on personal interests and is not as affected by the scholarly and academic environment. This dimension has two sub-dimensions. The Field sub-dimension refers to the extent to which the research agenda is influenced by scientific priorities that the field community determines by consensus (Becher & Trowler, 2001; Collins, 1994). A higher score for this sub-dimension means that the research agendas are more influenced by a community priority focus. The other sub-dimension, Institution, refers to the propensity of researchers to align their research agendas with the strategic research targets of their institutions. The higher the score for this sub-dimension, the greater this propensity will tend to be, while the lower the score, the greater the likelihood that the research agenda will be affected by institutional constraints. This propensity is expected to vary according to the sector in which the researcher is working (e.g., academia, industry, government, non-profit sector) and the career stage of the researcher, such that younger, untenured, and contract-based researchers will be more affected by institutional constraints (Giroux, 2015).

The second new dimension in the MDRAI-R is Society Driven, which measures the likelihood that a research agenda aims to solve challenges in society. The higher the score for this dimension, the greater the focus on such challenges, while the lower the score, the lesser the focus on such challenges. This dimension has two sub-dimensions. The first sub-dimension is Society, which refers to the incidence of society related challenges in a research agenda, while the second sub-dimension, Non-academics, measures the influence and participation of laymen and non-experts in the design of a research agenda. The higher the score for this sub-dimension, the greater the likelihood of engaging with non-research communities in an “action research community” or “participatory research” (Mendes et al., 2016; Wooltorton et al., 2015). These two sub-dimensions reflect the possibility of having a society-focused research agenda that does not involve collaboration with non-expert communities.

6. Conclusion

This study refines, extends, and optimizes the original MDRAI, which was validated only for the social sciences. Our revised MDRAI-R includes new dimensions and fewer items per dimension, and expands the scope and applicability of the inventory to all fields of knowledge. The new version exhibits good psychometric properties and satisfactory validity, reliability, and sensitivity. Furthermore, our measurement invariance analysis indicates that the model can be applied equally to all fields of knowledge, thus broadening its scope of application. The new dimensions (Academia Driven and Society Driven) provide new angles for assessing research agendas. This reinforces the usefulness of the instrument by allowing for cross-field studies while also identifying agendas with possible societal impact. Thus, in addition to being of interest to individual researchers, our instrument will be of value to policy makers, research funding agency strategists, and university and research organization leaders. In particular, the updated instrument will enable them to better characterize their research teams, and create incentives that can add value to their research. The final validated version is provided as an appendix to this study (Appendix B). The items are presented in no specific order and randomization is recommended before application to ensure that the gamification or fixed structuring of the questions does not result in biased responses.

This study has the following limitations. First, as with all perception-based measures, there is a risk of bias from the participants, and this possibility needs to be considered when reviewing the response data, especially with smaller samples. Second, the Academia Driven sub-scales are represented by only two items each. Although this is acceptable and not uncommon, it must be

noted that this is the absolute minimum number of items possible per factor. Thus, care should be taken when using the sub-scales alone rather than the overall Academia Driven measure, especially when data are missing. An additional limitation is that we could not test the external validity with current data. This is something we plan to address in future studies. Finally, some minor issues were identified with the convergent validity of the Discovery scale, and with the reliability of the Divergence scale. Although these are only decimal and millesimal deviations (respectively), they should still be noted, even though the practical impacts are likely to be negligible.

7. References

- Acker, S., & Webber, M. (2017). Made to measure: Early career academics in the Canadian university workplace. *Higher Education Research & Development*, 36(3), 541–554.
- Åkerlind, G., & McAlpine, L. (2017). Supervising doctoral students: Variation in purpose and pedagogy. *Studies in Higher Education*, 42(9), 1686–1698.
- Allison, P. D., Long, J. S., & Krauze, T. K. (1982). Cumulative advantage and inequality in science. *American Sociological Review*, 615–625.
- Anderson, D., Burnham, K., & White, G. (1998). Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies. *Journal of Applied Statistics*, 25(2), 263–282.
- Arbuckle, J. (2007). *Amos 16.0 user's guide*. Spss Chicago, IL.
- Bandalos, D. L. (1993). Factors influencing cross-validation of confirmatory factor analysis models. *Multivariate Behavioral Research*, 28(3), 351–374.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42(5), 815–824.
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. McGraw-Hill International.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238.

807 Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variables.
 808 *Psychometrika*, 45(3), 289–308.

809 Bentley, P. J., Gulbrandsen, M., & Kyvik, S. (2015). The relationship between basic and applied
 810 research in universities. *Higher Education*, 70(4), 689–709.

811 Bollen, K. A. (2014). *Structural equations with latent variables*. John Wiley & Sons.

812 Bourdieu, P. (1975). The specificity of the scientific field and the social conditions of the
 813 progress of reason. *Information (International Social Science Council)*, 14(6), 19–47.

814 Brew, A., Boud, D., Crawford, K., & Lucas, L. (2018). Navigating the demands of academic work
 815 to shape an academic job. *Studies in Higher Education*, 43(12), 2294–2304.

816 Brew, A., Boud, D., & Malfroy, J. (2017). The role of research education coordinators in building
 817 research cultures in doctoral education. *Higher Education Research & Development*,
 818 36(2), 255–268.

819 Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

820 Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus*
 821 *Editions*, 154, 136–136.

822 Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor
 823 covariance and mean structures: The issue of partial measurement invariance.
 824 *Psychological Bulletin*, 105(3), 456.

825 Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance.
 826 *Structural Equation Modeling*, 14(3), 464–504.

827 Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing
 828 measurement invariance. *Structural Equation Modeling*, 9(2), 233–255.

829 Ciarli, T., & Ràfols, I. (2018). The relation between research priorities and societal demands:
 830 The case of rice. *Research Policy*. <https://doi.org/10.1016/j.respol.2018.10.027>

831 Collins, R. (1994). *Why the social sciences won't become high-consensus, rapid-discovery*
 832 *science*. 9, 155–177. Springer.

833 DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores:
834 Considerations for the applied researcher. *Practical Assessment, Research &*
835 *Evaluation*, 14(20), 1–11.

836 Ebrahimi, S., & Osareh, F. (2014). Design, validation, and reliability determination a citing
837 conformity instrument at three levels: Normative, informational, and identification.
838 *Scientometrics*, 99(2), 581–597.

839 Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable
840 variables and measurement error. *Journal of Marketing Research*, 39–50.

841 Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research
842 strategies. *American Sociological Review*, 80(5), 875–908.

843 Franzoni, C., & Rossi-Lamastra, C. (2017). Academic tenure, risk-taking and the diversification
844 of scientific research. *Industry and Innovation*, 24(7), 691–712.

845 Gaskin, J. (2016). *Stats Tool Package*. Retrieved from <http://statwiki.kolobkcreations.com>

846 Giroux, H. A. (2015). *University in chains: Confronting the military-industrial-academic complex*.
847 Routledge.

848 Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2007). *Multivariate data*
849 *analysis*. Bookman.

850 Hancock, S., & Walsh, E. (2016). Beyond knowledge and skills: Rethinking the development of
851 professional identity during the STEM doctorate. *Studies in Higher Education*, 41(1),
852 37–50.

853 Hicks, D., & Katz, J. S. (2011). Equity and excellence in research funding. *Minerva*, 49(2), 137–
854 151.

855 Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for
856 determining model fit. *Articles*, 2.

857 Horta, H., & Santos, J. M. (2016). An instrument to measure individuals' research agenda
 858 setting: The multi-dimensional research agendas inventory. *Scientometrics*, 108(3),
 859 1243–1265. <https://doi.org/10.1007/s11192-016-2012-4>

860 Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis:
 861 Conventional criteria versus new alternatives. *Structural Equation Modeling: A*
 862 *Multidisciplinary Journal*, 6(1), 1–55.

863 Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced
 864 topics. *Journal of Consumer Psychology*, 20(1), 90–98.

865 Johnson, B., & Stevens, J. J. (2001). Exploratory and confirmatory factor analysis of the School
 866 Level Environment Questionnaire (SLEQ). *Learning Environments Research*, 4(3), 325.

867 Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A
 868 simulation study of categorization error. *American Sociological Review*, 398–407.

869 Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications*. Spss.

870 Jung, J. (2018). Learning experience and perceived competencies of doctoral students in Hong
 871 Kong. *Asia Pacific Education Review*, 42(4), 187–198.

872 Kenny, J. (2018). Re-empowering academics in a corporate culture: An exploration of workload
 873 and performativity in a university. *Higher Education*, 75(2), 365–380.

874 Kim, K., & Kim, J.-K. (2017). Inequality in the scientific community: The effects of cumulative
 875 advantage among social scientists and humanities scholars in Korea. *Higher Education*,
 876 73(1), 61–77.

877 Kingdon, M. (2013). *The science of serendipity: How to unlock the promise of innovation*. John
 878 Wiley & Sons.

879 Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford press.

880 Kuhn, T. S. (2012). *The structure of scientific revolutions*. University of Chicago press.

881 Kwiek, M. (2018). International research collaboration and international research orientation:
 882 Comparative findings about European academics. *Journal of Studies in International*
 883 *Education*, 22(2), 136–160.

884 Lanahan, L., Graddy-Reed, A., & Feldman, M. P. (2016). The domino effects of federal research
 885 funding. *PloS One*, 11(6), e0157325.

886 Latour, B., & Woolgar, S. (2013). *Laboratory life: The construction of scientific facts*. Princeton
 887 University Press.

888 Leahey, E. (2007). Not by productivity alone: How visibility and specialization contribute to
 889 academic earnings. *American Sociological Review*, 72(4), 533–561.

890 Luukkonen, T., & Thomas, D. A. (2016). The ‘negotiated space’ of university researchers’
 891 pursuit of a research agenda. *Minerva*, 54(1), 99–127.

892 Mantai, L. (2017). Feeling like a researcher: Experiences of early doctoral students in Australia.
 893 *Studies in Higher Education*, 42(4), 636–650.

894 Maroco, J. (2003). *Análise estatística: Com utilização do SPSS*.

895 Maroco, J. (2007). Consistency and efficiency of ordinary least squares, maximum likelihood,
 896 and three Type II linear regression models. *Methodology*, 3(2), 81–88.

897 Marôco, J. (2010). *Análise de equações estruturais: Fundamentos teóricos, software &*
 898 *aplicações*. ReportNumber, Lda.

899 McGrath, J. E. (1981). Dilemmatics: The study of research choices and dilemmas. *American*
 900 *Behavioral Scientist*, 25(2), 179–210.

901 Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit
 902 indices in tests of measurement invariance. *Journal of Applied Psychology*, 93(3), 568.

903 Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and
 904 confirmatory factor analytic methodologies for establishing measurement
 905 equivalence/invariance. *Organizational Research Methods*, 7(4), 361–388.

- 906 Mejia, C., & Kajikawa, Y. (2018). Using acknowledgement data to characterize funding
 907 organizations by the types of research sponsored: The case of robotics research.
 908 *Scientometrics*, 114(3), 883–904.
- 909 Mendes, R., Plaza, V., & Wallerstein, N. (2016). Sustainability and power in health promotion:
 910 Community-based participatory research in a reproductive health policy case study in
 911 New Mexico. *Global Health Promotion*, 23(1), 61–74.
- 912 Merton, R. K. (1968). The Matthew effect in science. *Science*, 159(3810), 56–63.
- 913 Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications
 914 in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–
 915 130.
- 916 Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances*
 917 *in Health Sciences Education*, 15(5), 625–632.
- 918 Ooms, W., Werker, C., & Hopp, C. (2018). Moving up the ladder: Heterogeneity influencing
 919 academic careers through research orientation, gender, and mentors. *Studies in Higher*
 920 *Education*, 1–22. <https://doi.org/10.1080/03075079.2018.1434617>
- 921 O’Rourke, N., & Hatcher, L. (2013). *A step-by-step approach to using SAS for factor analysis and*
 922 *structural equation modeling*. Sas Institute.
- 923 Polanyi, M. (2012). *Personal knowledge: Towards a post-critical philosophy*. University of
 924 Chicago Press.
- 925 Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting:
 926 The state of the art and future directions for psychological research. *Developmental*
 927 *Review*, 41, 71–90.
- 928 Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item
 929 short version of the Big Five Inventory in English and German. *Journal of Research in*
 930 *Personality*, 41(1), 203–212.

- 931 Roumbanis, L. (2018). Symbolic violence in academic life: A study on how junior scholars are
932 educated in the art of getting funded. *Minerva*, 1–22.
- 933 Santos, J. M., & Horta, H. (2018). The research agenda setting of higher education researchers.
934 *Higher Education*, 76(4), 649–668. <https://doi.org/10.1007/s10734-018-0230-9>
- 935 Schut, M., van Paassen, A., Leeuwis, C., & Klerkx, L. (2014). Towards dynamic research
936 configurations: A framework for reflection on the contribution of research to policy
937 and innovation processes. *Science and Public Policy*, 41(2), 207–218.
- 938 Shi, F., Foster, J. G., & Evans, J. A. (2015). Weaving the fabric of science: Dynamic network
939 models of science’s unfolding structure. *Social Networks*, 43, 73–85.
- 940 Shwed, U., & Bearman, P. S. (2010). The temporal structure of scientific consensus formation.
941 *American Sociological Review*, 75(6), 817–840.
- 942 Siciliano, M. D., Welch, E. W., & Feeney, M. K. (2018). Network exploration and exploitation:
943 Professional network churn and scientific production. *Social Networks*, 52, 167–179.
- 944 Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution
945 of sequential chi-square statistics. *Psychometrika*, 50(3), 253–263.
- 946 Sullivan, G. M., & Artino Jr, A. R. (2013). Analyzing and interpreting data from Likert-type
947 scales. *Journal of Graduate Medical Education*, 5(4), 541–542.
- 948 Wallace, M. L., & Ràfols, I. (2018). Institutional shaping of research priorities: A case study on
949 avian influenza. *Research Policy*, 47(10), 1975–1989.
- 950 Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford University
951 Press on Demand.
- 952 Whittaker, T. A. (2012). Using the modification index and standardized expected parameter
953 change for model modification. *The Journal of Experimental Education*, 80(1), 26–44.
- 954 Wooltorton, S., Wilkinson, A., Horwitz, P., Bahn, S., Redmond, J., & Dooley, J. (2015).
955 Sustainability and action research in universities: Towards knowledge for

956 organisational transformation. *International Journal of Sustainability in Higher*
957 *Education*, 16(4), 424–439.

958 Ying, Q. F., Venkatramanan, S., & Chiu, D. M. (2015). *Modeling and analysis of scholar mobility*
959 *on scientific landscape*. 609–614. ACM.

960 Young, M. (2015). Competitive funding, citation regimes, and the diminishment of
961 breakthrough research. *Higher Education*, 69(3), 421–434.

962 Zuckerman, H. (1978). Theory choice and problem choice in science. *Sociological Inquiry*, 48(3-
963 4), 65–95.

964 Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by
965 level of measurement? *Canadian Psychology/Psychologie Canadienne*, 34(4), 390.

966 Zuo, Z., & Zhao, K. (2018). The more multidisciplinary the better? The prevalence and
967 interdisciplinarity of research collaborations in multidisciplinary institutions. *Journal of*
968 *Informetrics*, 12(3), 736–756.

969