Are Exams Authentic Assessment? The Case of Economics

Petar Stankov[1], *Royal Holloway, University of London*

Exams have traditionally dominated assessment in higher education. Yet, we still do not know if they are authentic assessment, especially in an environment of rapid demographic changes within higher education and evolving perceptions of authenticity. This work offers a method to measure exam authenticity, defined as the overlap between exams and workplace assignments. To measure the overlap, workplace assignments and exams are mapped into a time-type space. The mapping creates a quantitative measure of exam authenticity. The results show that exams produce vital graduate job market skills. Therefore, they are authentic assessment and a necessary component of the post-pandemic assessment mix. The new authenticity measure is scalable and generalizable across fields and could inform policies and practices, such as subject benchmarking and student recruitment. The paper closes with a simple replication guide.

Keywords: authentic assessment; quantitative authenticity; post-COVID assessment; COVID-19

What is authenticity, and why is it important?

Assessment authenticity emerged in the work of (Wiggins, 1989), who defined it as a 'true test of intellectual ability' (p. 703). Such true test reflected the ability of a student to match certain intellectual standards but also replicated professional 'challenges and standards of performance' (Wiggins, 1989). Authentic assessment was then a concept used to test the knowledge deemed worth knowing within a certain field and was separated from the perceived monetary returns from that knowledge. In the broad version of the world outlined by Wiggins, testing workplace proficiencies was only one of the numerous elements of assessment authenticity (Newmann & Wehlage, 1993; Palm, 2008; Rennert-Ariev, 2005). Therefore, initially, assessment authenticity evolved a set of contexts, roles and metrics used to benchmark student performance (Frey et al., 2012) independently from their employment prospects. Exams or any other assessment did not need to necessarily evaluate a candidate's readiness for the job market. Instead, exams were primarily testing a set of abilities, competences and proficiencies students develop on their journey to a degree (Hansen, 1986), with the proficiency set being increasingly contextualised for the needs of a more diverse community of educators and students (Darling-Hammond & Snyder, 2000).

Practical Assessment, Research & Evaluation, Vol 29 No 3
Stankov, Are Exams Authentic Assessment?

Page 2

The links between workplace requirements for job-market candidates and assessment gained importance as higher education expanded into wider demographic groups. This narrowed the scope of what authentic means to be work-relevant assessment' (Ashford-Rowe et al., 2014), capable of replicating 'the tasks and performance standards typically found in the world of work' (Villarroel et al., 2018). More than 30 years since its inception, authenticity is broadly understood as a feature of assessment 'preparing the learner for what they are going to do next, meeting employer needs and testing knowledge and skills in a more realistic, contextualised and motivating way' (JISC, 2020).[2]

The conceptual evolution of authenticity happened because higher education itself evolved as an industry with a critical role for both individual and societal prosperity (Schultz, T. W., 1993; Woessmann, 2016). Growing demand and higher fees (Neill, 2009; Neill, 2015) refocused the industry consumers to the returns on their significant investment (Borooah & Mangan, 2008; Psacharopoulos & Patrinos, 2018). This was not specific to the fields of social sciences or business alone – perceived returns to education were driving demand across numerous fields of study (Jensen, 2010; Lemieux, 2014). It should be noted that classic works on the value of knowledge have long advocated developing skills for practical purposes, e.g. advancing individual, business, and political success (Spencer, 1860). However, subordinating university-level assessment to the needs of the world of work or inferring the value of a university degree from the lifetime income prospects of its graduates is a more recent approach.

Assessment in higher education is dominated by examinations. In economics, as in most other fields, an exam is a formal test of knowledge or ability in a particular subject (Collins Online Dictionary, 2023). Exams have been the backbone of assessment in general (Brown, 2022; Kellaghan & Greaney, 2019) and in economics in particular (Watts & Schaur, 2011) at least until the COVID-19 pandemic. The post-COVID assessment landscape, however, has been upended by both the shift to online assessment (Birdi et al., 2023) and the rise of artificial intelligence (Farazouli et al., 2023).

In this context, assessment innovation is not only natural but also necessary as the assessment mix is still skewed in favour of written exams. However, before educators shrink the role of exams in the post-COVID world, we need a robust debate on the following questions: Do exams lack authenticity? Will a reduction in the share of final exams necessarily improve authenticity? This paper is the first to address these sets of questions using a quantitative approach to authenticity, which is applicable to both economics and numerous fields beyond economics. This paper argues that a profound change in the share, format, and coverage of final exams in the post-COVID assessment landscape would be premature before educators develop a deeper understanding of exam authenticity in their fields. A method to gain this understanding is presented next.

Measuring exam authenticity

This approach outlined below fits the more contemporary, narrower, definition of assessment authenticity. This is done for practical purposes only, not because subordinating assessment to the needs of the job market is appropriate in all educational settings. The primary goal here is to create a measure of authenticity that is scalable across a variety of fields. However, scalability comes at a price. Adopting a classic, multi-dimensional definition of authenticity would trigger significant data collection challenges. For example, a meta-study of authenticity has identified nine dimensions (Frey et al., 2012): realism; performance orientation; complexity; defensibility; formativeness; collaborative nature; marking transparency; multiplicity of scoring indicators; and excellence. Mapping all nine for any single piece of assessment would pose insurmountable data challenges in most academic settings. In addition, many of the nine dimensions above are conceptually overlapping, and therefore collapsible to a smaller number.

---

[2] The conceptual evolution of authenticity has produced some exemplary classic and recent discussions, such as Petraglia (1998) and McArthur (2023).

Practical Assessment, Research & Evaluation, Vol 29 No 3
Stankov, Are Exams Authentic Assessment?

Page 3

Therefore, to limit the data requirements and benefit from the existing conceptual overlaps between the authenticity dimensions, we could collapse the dimensionality of how authenticity is defined. Almost none of the papers reviewed by Frey et al. (2012) discuss all nine dimensions, and many use between two and five dimensions.[3] This would also be in line with the more recent reductionist approach to authenticity, which brings the definition closer to the job market needs of graduates. For the purposes of this work, assessment authenticity is defined as the extent to which assessments mimic professional workplace assignments. Then, exams are more authentic whenever their overlap with workplace assignments grows.

Measuring the overlap between exams and workplace assignments requires screening of the skills demanded in the job market. Several large-scale studies inform of the existing and projected global demand for skills. For example, the World Economic Forum (WEF, 2023) ranks the 26 skills employers report as 'core' to their business. Among the top-5 core skills are cognitive skills, such as analytical and creative thinking, and self-efficacy skills, such as resilience, motivation, and curiosity. A variety of teamwork and technology skills complete the top-10 list (WEF, 2023, pp.37-48). At the same time, exams put precisely those analytical, creative, and self-efficacy skills to a test, unless they test a very basic, low-level, understanding of the subject and an ability to memorise content. The Organisation for Economic Cooperation and Development (OECD, 2017) has also studied the skills composition of the workforce in the digital age and has classified skills as 'low', 'medium-routine', 'medium non-routine' and 'high' (p.11), while academic literature has long used a triad of low-skilled, routine and analytical tasks to study the variety of skills input in the workplace (Autor & Handel, 2013). This means the types of exam problems, particularly in advanced modules students take before graduation, have a considerable overlap with the types of problems graduates face after graduation. It is therefore reasonable to design a measure of this overlap. The overlap uses three types of exam questions and three types of workplace assignments – routine, applied and creative – matching the types of skills tested at exams and demanded in the

workplace. It is this mix that typically shows up in an advanced exam and serves well in the workplace, according to both academic and policy-oriented literature.

The dominant methods to study authenticity have been discursive (Hagvall Svensson et al., 2022), thematic analyses of student and teaching staff self-reflections (Gulikers et al., 2004; Herrington & Oliver, 2000; Wiewiora & Kowalkiewicz, 2019), and data visualisation techniques (Schultz, M. et al., 2022). Meta-analyses have also emerged (Frey et al., 2012; Koh, 2017; Sokhanvar et al., 2021), with a particular contribution to the ways we define assessment authenticity.

Efforts to quantify authenticity followed soon after the concept was defined. Pioneering the quantitative field, Torrance (1995) offered a method to study teacher opinions on assessment authenticity. Nicaise et al. (2000) and Roach et al. (2018) added the student view but de-emphasized the perceptions of staff. The survey tradition was extended by Gulikers et al. (2008) and Kreber & Klampfleitner (2013), each of whom proposed a method to map current student and teacher perceptions into a measure of authenticity involving only the nature of assessment. Their link with the workplace assignments was rather indirect. In essence, due to its multi-dimensional definition and steep data requirements, the quantitative authenticity literature has so far focused on processing small self-reported datasets on *perceptions* of assessment authenticity. It has not built direct measures of the workplace applicability of college-level assessment. The current paper fills this gap within the survey tradition.

The closest work to this paper has been published by Schultz et al. (2022) and Schriebl et al. (2023). Schultz et al. (2022) adopt a job-market perspective on authenticity across four programmes in the STEM field. They obtain 199 student and 39 staff responses, which is a typical size of the sample in the field. They use data visualisation techniques to measure authenticity along several dimensions. Like Gulikers et al. (2008), they find significant similarities between the student and staff perceptions of authenticity. Schriebl et al. (2023) offer a multidimensional quantification of authenticity, contextualised for science education.

---

[3] For details, see the source table in Frey et al. (2012, p. 6)

However, neither quantifies the overlap between assessments and actual job market tasks by surveying both graduates and employers in ways that can be generalised across a variety of fields, which is an innovation of the current paper.

This paper advances the literature in four ways. First, it uses *alumni* – not current students – who are in a better position to evaluate the mapping between workplace assignments after graduation and assessments before graduation. Using current students to measure authenticity may bias the results. This is because current students typically have limited professional experience, and the resulting measure of authenticity would have reflected aspirations rather than workplace realities. Second, to strengthen the link with the job market, this paper features not only alumni but also actual employers. They add valuable perspectives on the nature and timing of workplace assignments college graduates receive in their first year on the job. Third, the views of academics, graduates and employers are studied using scalable methods to produce comparable measures of authenticity across departments or programmes within departments. Although the work here offers a snapshot of authenticity, its method is straightforward to extend over time, and guidance on it is offered in the Appendix. Fourth, this article measures authenticity at a time in a learners' development which has evaded scrutiny so far – the first year on the job after graduation. At this critical juncture in one's career, a fresh graduate still has memory of how exams were done and what was tested, but also has sufficient entry-level professional experience to map exams to professional workplace assignments. The survey methods to advance the quantitative authenticity literature are detailed below.

Methodology

The Time-Type Map

The overlap between exams and workplace assignments can be measured along several dimensions matching the chosen definition of authenticity. In this work, authenticity is measured along the following two dimensions: time and type. The time dimension measures a typical duration of an exam, and a typical duration of a job-market assignment. The duration of the exam is taken from the survey responses of academics, while the duration of the workplace assignment is taken from the responses of graduates and employers.

Most final assessments are typically given to students in a 3-hour window. Some of the more recent practice has extended the time frame of open-book exams to either 24 or 48 hours in a take-home setting (Burnett & Paredes Fuentes, 2020), or given students even more time if the final assignment is an essay or a report. As the 48-hour take-home exam is not significantly different from the 24-hour exam but allows for a more accurate overlap with non-routine workplace assignments, the 48-hour threshold was preferred over the 24-hour one.

Many fresh hires face similar timelines on their first jobs. Simple tasks are done within the day and, as tasks become more complicated and involve higher-order skills, graduates are given more time to complete them. Hence, it is natural to split time into three scales: less than 3 hours; between 3 and 48 hours; and more than 48 hours. The three time-scales are then used to measure the overlap between exams and workplace assignments along the time dimension.

The types of assessments effectively incorporate the nature of final exams and workplace assignments. Their type is also split along three values: routine, applied and creative. As those concepts may acquire different meanings in a variety of work and academic settings, the questionnaire ensured that academics, graduates, and employers fill them with similar meaning. Table 1 below presents the definitions of the types of assessment and workplace assignments used in the questionnaire. At the same time, the Appendix presents the full questionnaire and offers further guidance for replicating the methodology. This is done in the hope that educators will be able to quantify authenticity in their own departments.

The split between routine, applied and creative tasks serves three purposes. First, the split ensures that most types of exam questions and workplace assignments of fresh graduates are captured by the questionnaire. Second, the split is rooted in well-established pedagogy where low-level skills (e.g., knowledge and understanding) are typically routine; mid-level skills (e.g., application and analysis) are typically where students or graduates use basic skills in novel scenarios; and high-level skills (e.g. synthesis, evaluation, decision making) are the ones where

Practical Assessment, Research & Evaluation, Vol 29 No 3
Stankov, Are Exams Authentic Assessment?

Page 5

creativity matters most (Bloom et al., 1956; Dubas & Toledo, 2016; Krathwohl, 2002). Third, constraining the assessment typology to just three-by-three prevents a data-hungry proliferation beyond what is feasible to capture in a small survey once the time and type dimensions are combined.

Combining the three time-scales and the three types of assessments and assignments gives a three-by-three map. Once both final exams and workplace assignments are categorised into each of the map cells, calculating the overlap between them is straightforward using the data and methods described further.

Data

Survey responses were collected in line with ethical and data protection standards in May-June 2022 from 144 academics, graduates, and employers. Respondents included most staff members of an economics department, as well as the significant alumni network of the same department. To boost participation from regions outside of the UK, the survey was cascaded to economics educators using personal contacts. The resulting sample size is standard in the field. For example, the number of observations in surveys using current student responses ranges from 46 (Kreber & Klampfleitner, 2013) to 305 (Roach et al., 2018). Studies involving academics exhibit similar sample sizes, ranging from 59 (Nicaise et al., 2000) to 199 (Schultz, M. et al., 2022).

In this study, respondents either teach economics in higher education, apply economics in their work life soon after graduation from an economics department, or are economists who hire or supervise other fresh economics graduates. Limiting the survey sample to economists was necessary to ensure that respondents had similar understanding of the concepts used in the questionnaire. Although the sample is limited, the questionnaire itself is general enough to be used across fields. In turn, any overlap in the 3x3 map would not be accidental or driven by a statistical anomaly. In addition, limiting the duration of the study to several weeks was necessary to ensure respondents find themselves in a similar phase of the business cycle. The survey was anonymous, and no personal data was collected unless the respondent opted in to hear back about the results by entering an email address. Respondents granted permission to use their answers for the purposes of this study.

Responses came back from 55 academics, 19 employers, and 70 economics graduates. Like any self-reported data, perceptions of authenticity by the respondents may impose some measurement errors. The questionnaire was designed to minimise the risk of subjectivity by asking simple questions on the duration and type of exams and assignments.[4] To get a feel of the data, the figure below presents the type of a typical question on an exam and workplace assignment, as well as their time to completion.

Table 1. Definitions of Assessment and Workplace Assignment Types

Academic	Workplace
The typical question on my exam (final assessment) is: - **Routine**: It tests the ability of my students to understand the material (know the main facts, theories, discipline-specific formulas) that they have been exposed to numerous times during the term. - **Applied**: It tests the ability of my students to apply the material to known or novel situations or scenarios, including contemporary events. - **Creative**: It tests the ability of my students to critically evaluate the discipline knowledge using existing or novel techniques and, on occasion, create new knowledge.	The typical assignment our new employees receive are: - **Routine**: related to reproducing basic knowledge about the job or the field their degree is in. - **Applied**: related to the application of basic or more advanced knowledge in novel situations or scenarios. - **Creative**: related to a critical evaluation of existing practices; suggesting novel processes, practices, or products; leading project teams.

---

[4] See the Appendix for details.

Figure 1. Type and Duration of Exams and Workplace Assignments
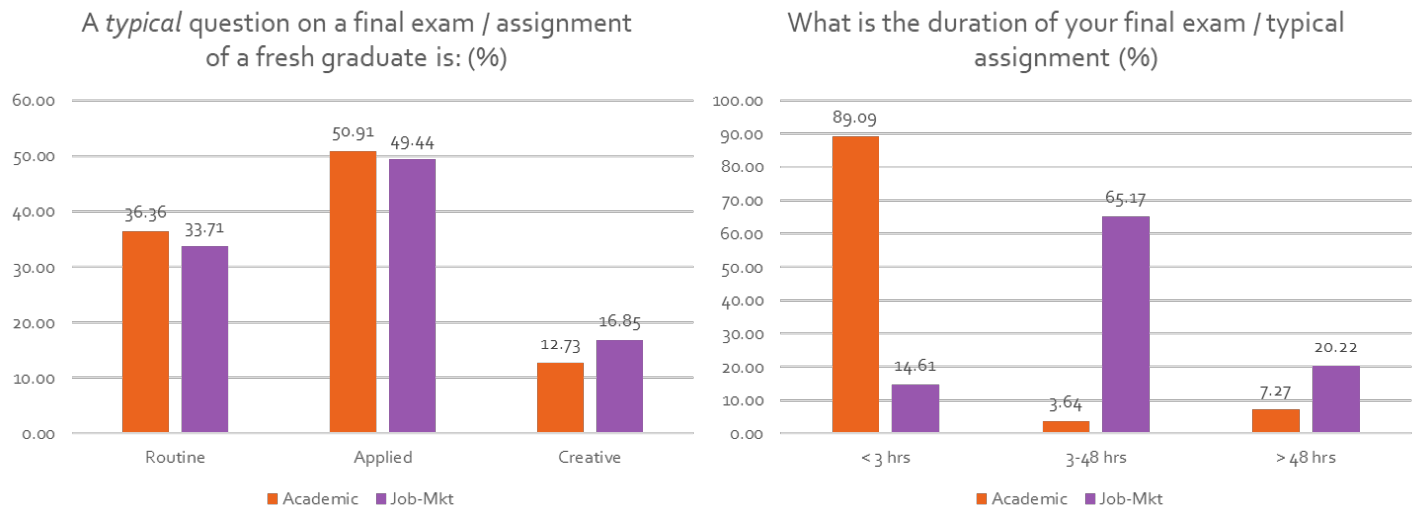
The above responses allow for two initial observations. First, the type of typical exam questions and workplace assignments overlap to a great extent. Second, students are more pressed for time to complete their final exams than they are in the workplace. These observations set the stage for the main question of this work: What is the overlap between exams and workplace assignments? The methods below outline how this question was addressed.

Methods

To answer the main question of this work, two sets of methods were adopted. First, I created heatmaps following the emerging quantitative literature on assessment authenticity (Schultz, M. et al., 2022). The heatmaps cover the two dimensions of authenticity in this study: time and type. However, heatmaps cannot test if the differences between the shares of exams and job market assignments are statistically significant. This is where the current paper delivers its innovation: It produces a measure of authenticity and summarises it in an authenticity index.

To arrive at the authenticity index, I first run bivariate regressions of a continuous share of routine, applied or creative questions on a dummy variable equal to one if the respondent belongs to the workplace. Any difference in the parameters of interest represents divergence between exams and workplace assignments. This difference, D_{ij}, is the slope coefficient in the bivariate regression in each of the three time-type rows indicated by i and each of the three columns indicated by j. For example, cell D_{11} compares the share of routine questions on exams with a duration shorter than 3 hours to the share of routine workplace tasks that need completion within 3 hours, and stores the difference. Similarly, cell D_{22} stores the difference for applied questions needing solutions within 48 hours for both exams and workplace assignments, and cell D_{33} includes creative questions demanding solutions in more than 48 hours. The method to produce D_{11} is then replicated to produce D_{ij} for each cell in the three-by-three map.

Then, I average D_{ij} across the entire map using a simple mean of the differences: $D = \frac{1}{C} \sum_j \sum_i D_{ij}$, where C is the number of cells in the map. As each $D_{ij} \in [-1; 1]$, the overall authenticity index D will also vary between -1 and 1.

It is worth noting that the dominant location of most respondents is the UK (73% of graduates, 51% of academics and 47% of employers). Therefore, the results presented below will inevitably reflect a bias towards the authenticity of final exams run in the UK. The following section discusses the results.

Figure 2. Time-Type Heatmaps for Final Exams and Workplace Assignments
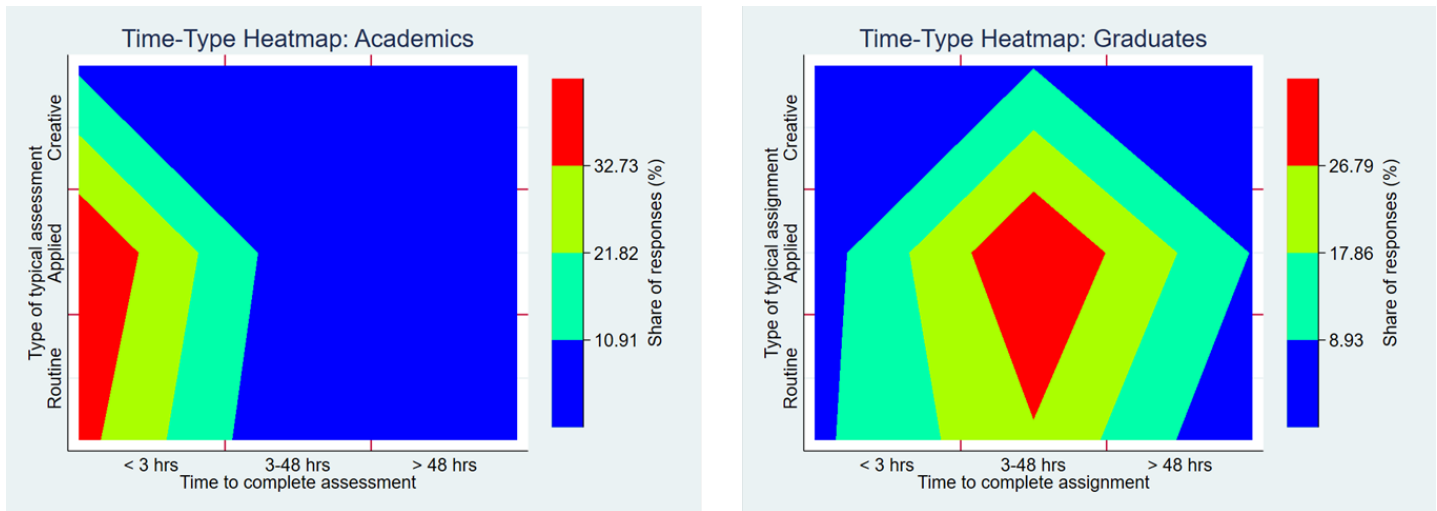
Results

The methodology above produces two sets of results: heatmaps and regression output. The heatmaps reveal what was already suspected from the preliminary data observations: Final exams prepare job market candidates for a broad range of frequently seen workplace assignments.

The types of typical final exam questions are mostly routine or applied, and very few of the typical questions (the ones that students see most often) on a final exam are creative. To some extent, this is normal as giving mostly creative questions on a final exam is riskier for an academic – that exam carries a higher likelihood of generating a skewed distribution of marks. However, the same holds for workplace assignments. Employers rarely give tasks to fresh hires that would disrupt their own products or processes. Most new hires work on routine or applied tasks, just like they would do on a typical final exam.

The differences between final exams and workplace assignments are more evident in the time to completion. Due to prevalent academic tradition, most routine final exams are given within a 3-hour window. However, perhaps due to the complexity of workplace assignments which typically involve teamwork, a workplace assignment of routine or applied nature would typically be due in 48 hours, not just 3. It appears that fresh graduates are endowed with more time to complete an assignment than students in an exam setting.

To some extent this is desirable as, from an academic perspective, one wants a fresh graduate to be able to easily hit a deadline in the world beyond the classroom. Similarly, from a job market perspective, one wants to hire graduates ready to submit a routine or applied assignment within a reasonably short deadline. The alternative world – one where academics produce assessments with too much embedded slack, while graduates struggle to meet deadlines in their first year on the job – would be undesirable.

There is only so much information that can be distilled form the heatmaps alone. We now turn to the regression output which sheds further light on the overlap between final exams and workplace assignments of fresh graduates. The results are presented in Table 2, Table 3, and Table A-1.

Table 2 has nine cells, each corresponding to a bivariate regression as detailed in the previous section. Because the regressions are bivariate with no additional controls, the name of the resulting index – simple bivariate authenticity – carries over. The simple bivariate authenticity produced by averaging all slope coefficients for the regressions with sufficient number of observations is $D_1 = -0.055$. Recall that the lower the difference between exams and assignments, and the lower the statistical significance of the difference, the higher is their overlap. Therefore, an overall authenticity index closer to zero will indicate more authentic assessment. As the simple bivariate authenticity index D_1 is close to zero, it suggests that the problems students face in their final exams are

Practical Assessment, Research & Evaluation, Vol 29 No 3
Stankov, Are Exams Authentic Assessment?

Page 8

insignificantly different from the ones students receive in their first year on the job. In short, when using the above measure, we can show that economics exams are authentic assessment.

Still, bivariate regressions are rather naïve without additional controls, such as the locations of respondents. Including the location would control for the fact that the exam structure, coverage, and difficulty, as well as workplace assignments, may differ across geographies. Table 3 presents the results of the same regressions as above with added location characteristics.

Table 3 tells a similar story on assessment authenticity. Conditioning authenticity on the location of the respondents does not alter the main result: The differences between exams and workplace assignments are small and insignificant. With an overall adjusted (for location) authenticity index of $D_2 = -0.069$, the results look comforting.

Table 2. Simple Bivariate Authenticity

Creative	-.183	No obs.	.125
	(.101)		(.134)
Applied	-.073**	-.108	.000
	(.032)	(.098)	(.061)
Routine	-.092	No obs.	No obs.
	(.065)		
	< 3 hrs	3-48 hrs	> 48 hrs

Notes: Results are from regressing the share of Routine, Applied or Creative questions on indicators that respondents are employers or economics graduates. The dependent variable is the share of Routine questions / assignments in cells 1, 2 and 3; the share of Applied questions / assignments in cells 4, 5 and 6; and the share of Creative questions / assignments in cells 7, 8 and 9. Cell 1 is in the lower-left, while cell 9 is in the upper-right corner of the table. Robust standard errors are presented in parentheses. Symbols: * p < 0.1; ** p < 0.05; *** p < 0.01; No obs. signals insufficient observations to run the regressions.

Table 3. Simple Bivariate Authenticity, Conditional on Location

Creative	-.205	No obs.	.063
	(.133)		(.137)
Applied	-.078	-.063	-.036
	(.046)	(.148)	(.032)
Routine	-.092	No obs.	No obs.
	(.074)		
	< 3 hrs	3-48 hrs	> 48 hrs

Notes: Results are from regressing the share of Routine, Applied or Creative questions on indicators that respondents are employers or economics graduates, and location dummies for the UK, the USA, the EU, other Europe, and Asia. The dependent variable is the share of Routine questions / assignments in cells 1, 2 and 3; the share of Applied questions / assignments in cells 4, 5 and 6; and the share of Creative questions / assignments in cells 7, 8 and 9. Robust standard errors are presented in parentheses. Symbols: * p < 0.1; ** p < 0.05; *** p < 0.01; No obs. signals insufficient observations to run the regressions.

Practical Assessment, Research & Evaluation, Vol 29 No 3
Stankov, Are Exams Authentic Assessment?

Page 9

Finding that economics exams are authentic assessment is indeed encouraging. However, three out of the nine cells in Table 2 and Table 3 do not possess enough observations to run a bivariate regression. To some extent this is expected, as most final exams are limited within a 3-hour window. Seeing exams with typically routine questions run for more than 48 hours would indeed be an academic oddity.

To ensure enough observations within types, I bunch together the time scales. Rather than run nine regressions, I now run only three – one for each type of exam question and workplace assignment. The results are presented in Table A-1 in the Appendix. As seen from the table, the main message still stands, with both the simple and the adjusted authenticity index hovering around zero, we can derive some degree of comfort about economics exams: they are similar in nature to workplace assignments. The following section discusses the results before a conclusion.

Discussion and conclusions

Despite classic criticisms on their capacity to measure student proficiencies (Becker, 1982; Cox, 1967), exams have withstood the test of time and dominated the assessment landscape at least until the COVID-19 pandemic. The post-COVID assessment landscape, however, has been upended by both the shift to online assessment, which is prone to cheating, collusion, and commissioning (Birdi et al., 2023) and the rise of artificial intelligence (Farazouli et al., 2023). Like previous waves of automation, artificial intelligence has triggered a tide of predictable changes in the workplace. Among the key changes is the reshuffling of routine, applied and creative tasks between machines and humans (Autor, 2015). In this rapidly evolving landscape, we need measures of assessment authenticity that stand the test of time and can be applied to both old and new assessments across a variety of fields. Such measure would either offer reassurance that exams offer adequate transferrable skills for the entry-level jobs that students want or motivate an assessment reform to heighten authenticity.

This paper has offered an approach to measuring exam authenticity. Efforts to quantify authenticity exist and have been only gaining traction. However, they still do not offer methods to map assessments to actual job market tasks in ways that can be generalised across a variety of fields. Such quantitative maps are necessary for a variety of reasons.

First, despite their positive impact on the formation of job-related skills (Leschnig et al., 2022), exams perpetuate attainment gaps across ethnic (Shaw & Tranter, 2021), racial (Sohn, 2012) and gender (Ahlburg & McCall, 2021) groups. Attainment gaps (AGs) will then go on to cement social inequalities (Baert & Verhaest, 2021; Feng & Graetz, 2017; Tampieri, 2016). Therefore, both a government interested in reducing attainment inequalities and a university accepting a high share of students from diverse backgrounds will push for a lower share of final exams at the programme level. Indeed, UK government documents suggest that a policy window for such reduction is now open (Hobbs & Mutebi, 2021). Using the known benefits of programme-level design (Myers et al., 2009; Salemi & Siegfried, 1999), post-pandemic assessment reforms will likely reach far and wide across higher education. As final exams have traditionally occupied the highest share of assessment while attracting vocal criticisms, they would become natural candidates for assessment reform. However, to help those reforms, we need methods to evaluate the authenticity of both exams and their replacements. This paper has proposed such a method, which is scalable across fields and informative of the overlap between the content and contexts students face in both exams and their entry-level jobs.

This overlap may not be the only measure or even the single most important measure of authenticity as some skills may gain importance over time even without a particular currency in the job-market. However, as higher education enters increasingly diverse populations of students who are more career-minded and sensitive to upward mobility prospects after graduation (Longwell-Grice et al., 2016; Pascarella et al., 2004; Terenzini et al., 1996), measuring the overlap offers a way to justify both existing and novel assessments.

Second, to build intra-pandemic assessment resilience, economics and other social science departments across the globe have already boosted the share of continuous assessment at the expense of final exams. This shift was natural as spreading assessment over time within an academic year mitigated the risk of severe disruptions to high-stake exams (Stankov,

Practical Assessment, Research & Evaluation, Vol 29 No 3
Stankov, Are Exams Authentic Assessment?

Page 10

2023). The emergency assessment transformation has largely served its purpose as testing the learning outcomes can arguably be achieved with a significant relaxation of exam constraints or, in some cases, even without final exams (Burnett & Paredes Fuentes, 2020; Villarroel et al., 2020). However, we do not yet know if the novel assessment methods would be perceived by either students or educators as authentic until we can adopt a framework for measuring authenticity.

Third, a measure of authenticity that is scalable across fields of study adds a key performance indicator to benchmarking exercises across departments and universities (Tauer et al., 2007; Worthington, 2001). As a result, the authenticity of final exams could inform the work done at the sectoral level to improve the quality of higher education, i.e., the subject benchmark statements by the Quality Assurance Agency for Higher Education (QAA) and other equivalent bodies.

Fourth, once a quantitative measure of authenticity exists, departments could use it for external promotion. For example, they could credibly argue that they place employability at the heart of their assessment strategy. Alternatively, they could adjust their assessment models for authenticity to facilitate programme-level assessment development and add new dimensions to their assessment (Walstad, 2001).

The approach to quantifying authenticity involved the creation of a two-dimensional time-type space. This enabled an evaluation of the overlaps between exams and content and conditions students encounter after graduation. This overlap has conceptual roots in earlier studies of authenticity (Palm, 2008) and is measured using survey data, which is standard in the field. Simple aggregation then helped create the authenticity indices. The indices allow for the conclusion that exams mimic workplace practices and therefore still produce vital workplace skills for economics graduates. Extensions to other fields using the methodology suggested here would be straightforward.

The emphasis given in this work on the workplace applicability of higher education knowledge, skills and processes is deliberate. In its more recent incarnations, assessment authenticity maps the demand for skills in the job market onto key metrics of student proficiencies. This mapping is implicit not only in economics and social sciences, but also across other fields through the perceived returns to education,

which allocate demand for education across fields. Therefore, the views we form about assessment authenticity in economics gain relevance, particularly for fields where the assessment landscape is still dominated by exams. Even though the perspectives of employers and recent graduates may not be exhaustive of the ways higher education enriches a graduate, their views were key to gauge the overlap between exams and workplace assignments. Admittedly, the approach taken here is somewhat reductionist as it has limited the conceptual space of authenticity in exchange for observational convenience. Future research will perhaps relax some of these constraints to gain further valuable insights into quantitative assessment authenticity.

The results in this paper call for exercising caution when abolishing exams. A careful assessment reform geared towards assessment authenticity will mandate an *ex-ante* authenticity study. This paper has offered a straightforward way to conduct such study within any field where assessment is dominated by exams. The Appendix presents replication guidance.

Statements and Declarations

Conflicting interests: The author does not have competing financial or non-financial interests that are directly or indirectly related to the work submitted for publication. No funding was received for conducting this study.

Compliance with Ethical Standards: Human respondents filled questionnaires to collect the data used in this work with their informed consent. No human trials were involved in collecting this data, and participants were not subject to any treatment other than recollecting facts from their own academic or professional experiences. Hence, an approval by an ethics committee was not applicable. No data has been collected or will be disclosed that could in any way reveal the anonymity of any of the human participants in this study.

References

Ahlburg, D. A., & McCall, B. P. (2021). One hundred years of the gender gap in examination results at the university of oxford. *History of Education, 50*(2), 240-256. 10.1080/0046760X.2020.1836266

Practical Assessment, Research & Evaluation, Vol 29 No 3
Stankov, Are Exams Authentic Assessment?

Page 11

Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education, 39*(2), 205-222. 10.1080/02602938.2013.819566

Autor, D. H. (2015). Why are there still so many jobs? the history and future of workplace automation. *Journal of Economic Perspectives, 29*(3), 3-30. 10.1257/jep.29.3.3

Autor, D. H., & Handel, M. J. (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics, 31*, S59-S96. 10.1086/669332

Baert, S., & Verhaest, D. (2021). Work hard or play hard? degree class, student leadership and employment opportunities. *Oxford Bulletin of Economics and Statistics, 83*(4), 1024-1047. 10.1111/obes.12424

Becker, W. E. (1982). The educational process and student achievement given uncertainty in measurement. *The American Economic Review, 72*(1), 229-236. http://www.jstor.org/stable/1808591

Birdi, A., Cook, S., Elliott, C., Lait, A., Mehari, T., & Wood, M. (2023). A critical review of recent economics pedagogy literature, 2020–2021. *International Review of Economics Education, 43*, 100264. 10.1016/j.iree.2023.100264

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. handbook 1: Cognitive domain.* D. McKay.

Borooah, V. K., & Mangan, J. (2008). Education, occupational class, and unemployment in the regions of the United Kingdom. *Education Economics, 16*(4), 351-370. 10.1080/09645290801977157

Brown, G. T. L. (2022). The past, present and future of educational assessment: A transdisciplinary perspective. *Frontiers in Education, 7*10.3389/feduc.2022.1060633

Burnett, T., & Paredes Fuentes, S. (2020). *Assessment in the time of pandemic: A panic-free guide.* Bristol, UK: The Economics Network. 10.53593/n3298a

Collins Online Dictionary. (2023). *Examination.* Collins Online Dictionary. Retrieved 06 October 2023, from https://www.collinsdictionary.com/dictionary/english/examination

Cox, R. (1967). Examinations and higher education: A survey of the literature. *Higher Education Quarterly, 21*(3), 292-340. 10.1111/j.1468-2273.1967.tb00244.x

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education, 16*(5-6), 523-545. 10.1016/S0742-051X(00)00015-9

Dubas, J. M., & Toledo, S. A. (2016). Taking higher order thinking seriously: Using Marzano's taxonomy in the economics classroom. *International Review of Economics Education, 21*, 12-20. 10.1016/j.iree.2015.10.005

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., & McGrath, C. (2023). Hello GPT! Goodbye home examination? An exploratory study of AI chatbots impact on university teachers' assessment practices. *Assessment & Evaluation in Higher Education,* 1-13. 10.1080/02602938.2023.2241676

Feng, A., & Graetz, G. (2017). A question of degree: The effects of degree class on labor market outcomes. *Economics of Education Review, 61*, 140-161. 10.1016/j.econedurev.2017.07.003

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical Assessment, Research, and Evaluation, 17*(2) https://doi.org/10.7275/sxbs-0829

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development, 52*(3), 67. 10.1007/BF02504676

Gulikers, J. T. M., Bastiaens, T. J., Kirschner, P. A., & Kester, L. (2008). Authenticity is in the eye of the beholder: Student and teacher perceptions of assessment authenticity. *Journal of Vocational Education & Training, 60*(4), 401-412. 10.1080/13636820802591830

Hagvall Svensson, O., Adawi, T., & Johansson, A. (2022). Authenticity work in higher education learning environments: A double-edged sword? *Higher Education, 84*, 67-84. 10.1007/s10734-021-00753-0

Practical Assessment, Research & Evaluation, Vol 29 No 3
Stankov, Are Exams Authentic Assessment?

Page 12

Hansen, W. L. (1986). What knowledge is most worth knowing--for economics majors? *The American Economic Review, 76*(2), 149-152. http://www.jstor.org/stable/1818754

Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development, 48*(3), 23-48. 10.1007/BF02319856

Hobbs, A., & Mutebi, N. (2021). *Inequalities in education, and attainment gaps*. London, UK: UK Parliament. https://post.parliament.uk/inequalities-in-education-and-attainment-gaps/

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics, 125*(2), 515-548. http://www.jstor.org/stable/27867489

JISC. (2020). *The future of assessment: Five principles, five targets for 2025*. https://www.jisc.ac.uk/reports/the-future-of-assessment

Kellaghan, T., & Greaney, V. (2019). A brief history of written examinations. In T. Kellaghan, & V. Greaney (Eds.), *Public examinations examined* (pp. 350). The World Bank. 10.1596/978-1-4648-1418-1_ch3

Koh, K. H. (2017). *Authentic assessment*. Oxford University Press. 10.1093/acrefore/9780190264093.013.22

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212-218. 10.1207/s15430421tip4104_2

Kreber, C., & Klampfleitner, M. (2013). Lecturers' and students' conceptions of authenticity in teaching and actual teacher actions and attributes students perceive as helpful. *Higher Education, 66*(4), 463-487. 10.1007/s10734-013-9616-x

Lemieux, T. (2014). Occupations, fields of study and returns to education. *The Canadian Journal of Economics / Revue Canadienne D'Economique, 47*(4), 1047-1077. http://www.jstor.org/stable/43818764

Leschnig, L., Schwerdt, G., & Zigova, K. (2022). Central exams and adult skills: Evidence from PIAAC. *Economics of Education Review, 90*, 102289. 10.1016/j.econedurev.2022.102289

Longwell-Grice, R., Adsitt, N. Z., Mullins, K., & Serrata, W. (2016). The first ones: Three studies on first-generation college students. *NACADA Journal, 36*(2), 34-46. 10.12930/NACADA-13-028

McArthur, J. (2023). Rethinking authentic assessment: Work, well-being, and society. *Higher Education, 85*(1), 85-101. 10.1007/s10734-022-00822-y

Myers, S. C., Nelson, M. A., & Stratton, R. W. (2009). Assessing an economics programme: Hansen proficiencies, ePortfolio, and undergraduate research. *International Review of Economics Education, 8*(1), 87-105. 10.1016/S1477-3880(15)30077-3

Neill, C. (2009). Tuition fees and the demand for university places. *Economics of Education Review, 28*(5), 561-570. 10.1016/j.econedurev.2009.01.002

Neill, C. (2015). Rising student employment: The role of tuition fees. *Education Economics, 23*(1), 101-121. 10.1080/09645292.2013.818104

Newmann, F., & Wehlage, G. (1993). Five standards of authentic instruction. *Educational Leadership, 50*(7)

Nicaise, M., Gibney, T., & Crane, M. (2000). Toward an understanding of authentic learning: Student perceptions of an authentic classroom. *Journal of Science Education and Technology, 9*(1), 79-94. 10.1023/A:1009477008671

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

Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research, and Evaluation, 13*(4)

Pascarella, E. T., Pierson, C. T., Wolniak, G. C., & Terenzini, P. T. (2004). First-generation college students. *The Journal of Higher Education (Columbus), 75*(3), 249-284. 10.1080/00221546.2004.11772256

Petraglia, J. (1998). *Reality by design. the rhetoric and technology of authenticity in education*. Routledge. 10.4324/9781410601254

Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: A decennial review of the global literature. *Education Economics, 26*(5), 445-458. 10.1080/09645292.2018.1484426

Practical Assessment, Research & Evaluation, Vol 29 No 3
Stankov, Are Exams Authentic Assessment?

Page 13

Rennert-Ariev, P. (2005). A theoretical model for the authentic assessment of teaching. *Practical Assessment, Research & Evaluation, 10*(1), 2. 10.7275/a7h7-4111

Roach, K., Tilley, E., & Mitchell, J. (2018). How authentic does authentic learning have to be? *Higher Education Pedagogies, 3*(1), 495-509. 10.1080/23752696.2018.1462099

Salemi, M. K., & Siegfried, J. J. (1999). The state of economic education. *The American Economic Review, 89*(2), 355-361. http://www.jstor.org/stable/117135

Schriebl, D., Müller, A., & Robin, N. (2023). Modelling authenticity in science education. *Science & Education, 32*(4), 1021-1048. 10.1007/s11191-022-00355-x

Schultz, M., Young, K., Gunning, T. K., & Harvey, M. L. (2022). Defining and measuring authentic assessment: A case study in the context of tertiary science. *Assessment & Evaluation in Higher Education, 47*(1), 77-94. 10.1080/02602938.2021.1887811

Schultz, T. W. (1993). The economic importance of human capital in modernization. *Education Economics, 1*(1), 13-19. 10.1080/09645299300000003

Shaw, L. M., & Tranter, M. R. (2021). Levelling the playing field in assessment: An analysis of attainment gaps for widening participation, Black and Minority Ethnic mathematics undergraduates before and after the COVID-19 lockdown. *Teaching Mathematics and its Applications: An International Journal of the IMA, 40*(4), 497-505. 10.1093/teamat/hrab024

Sohn, K. (2012). The dynamics of the evolution of the Black–White test score gap. *Education Economics, 20*(2), 175-188. 10.1080/09645292.2010.488487

Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation, 70*, 101030. 10.1016/j.stueduc.2021.101030

Spencer, H. (1860). What knowledge is of most worth? *Education: Intellectual, moral, and physical.* (pp. 21-96). D Appleton & Company. 10.1037/12158-001

Stankov, P. (2023). *Is hy-flex teaching sustainable post-COVID? Insights from an economics department.* Bristol, UK: The Economics Network. 10.53593/n3592a https://economicsnetwork.ac.uk/showcase/stankov_hyflex

Tampieri, A. (2016). Social background effects on school and job opportunities. *Education Economics, 24*(5), 496-510. 10.1080/09645292.2015.1074983

Tauer, L. W., Fried, H. O., & Fry, W. E. (2007). Measuring efficiencies of academic departments within a college. *Education Economics, 15*(4), 473-489. 10.1080/09645290701419516

Terenzini, P. T., Springer, L., Yaeger, P. M., Pascarella, E. T., & Nora, A. (1996). First-generation college students: Characteristics, experiences, and cognitive development. *Research in Higher Education, 37*(1), 1-22. 10.1007/BF01680039

Torrance, H. (1995). *Evaluating authentic assessment: Problems and possibilities in new approaches to assessment.* Open University Press.

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education, 43*(5), 840-854. 10.1080/02602938.2017.1412396

Villarroel, V., Boud, D., Bloxham, S., Bruna, D., & Bruna, C. (2020). Using principles of authentic assessment to redesign written examinations and tests. *Innovations in Education and Teaching International, 57*(1), 38-49. 10.1080/14703297.2018.1564882

Walstad, W. B. (2001). Improving assessment in university economics. *The Journal of Economic Education, 32*(3), 281-294. 10.2307/1183385

Watts, M., & Schaur, G. (2011). Teaching and assessment methods in undergraduate economics: A fourth national quinquennial survey. *The Journal of Economic Education, 42*(3), 294-309. http://www.jstor.org/stable/23049285

WEF. (2023). *The future of jobs report 2023.* World Economic Forum. https://www.weforum.org/reports/the-future-of-jobs-report-2023

Wiewiora, A., & Kowalkiewicz, A. (2019). The role of authentic assessment in developing authentic

leadership identity and competencies. *Assessment & Evaluation in Higher Education, 44*(3), 415-430. 10.1080/02602938.2018.1516730

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *The Phi Delta Kappan, 70*(9), 703-713. http://www.jstor.org/stable/20404004

Woessmann, L. (2016). The economic case for education. *Education Economics, 24*(1), 3-32. 10.1080/09645292.2015.1059801

Worthington, A. C. (2001). An empirical survey of frontier efficiency measurement techniques in education. *Education Economics, 9*(3), 245-268. 10.1080/09645290110086126

Corresponding Author:
Petar Stankov
Economics Department, Royal Holloway, University of London

Email: Petar.Stankov [at] rhul.ac.uk

Appendix

A1. Brief guidance on local replication
The following guidance will help you apply the methods here in your own context. It would be best to collect data on multiple courses for a programmatic evaluation of authenticity. First, ask colleagues in your department to fill out the questionnaire below. Second, get in touch with as many recent alumni as you can. To find alumni from your own department, use your university alumni office and/or connect with them on LinkedIn or other professional social media platform. Most graduates from your department will have a LinkedIn profile even if they are not registered with the alumni office. Third, your department's professional network on LinkedIn will feature employers. Get in touch with them. Some older alumni will have already assumed high-level executive positions or involved in hiring for their companies. Ask them to fill out the questionnaire as employers. Finally, you can use any survey collection platform to collect all responses. The survey responses below were collected using Microsoft Forms, which allows a direct data export to Excel.

A2. Questionnaire

Section 1: Branching question
1.Are you an: (Required to answer. Single choice.)
- Employer
- College Graduate (New Employee)
- Academic

[The questionnaire then branches out into relevant sections.]

Section 2: Employer survey
2. A fresh out-of-college hire in my company is employed at a: (Required to answer. Single choice.)
- temporary position (e.g. internship) with a fixed-term contract.
- full-time position where the new hire enjoys an open-ended contract after their probationary term.

3. We know that fresh graduates are involved in performing a variety of tasks, assignments, and projects. Some of them are routine, while others are creative. Some take just a few hours to complete, while others may last months. However, all new hires in your organisation are involved in some types of assignments more often than others. Let's call those typical assignments: the ones new hires receive most often.
The typical assignment your new employees receive is: (Required to answer. Single choice.)
- Routine: related to reproducing basic knowledge about the job or the field their degree is in.
- Applied: related to application of basic or more advanced knowledge in novel situations or scenarios.
- Creative: related to a critical evaluation of existing practices; suggesting novel processes, practices, or products; leading project teams.

4. The approximate share of routine tasks a new hire receives in their first year with us is: (Required to answer. Single line text.) Enter your answer.

5. The approximate share of applied tasks a new hire receives in their first year with us is: (Required to answer. Single line text.) Enter your answer.

6. The approximate share of creative tasks a new hire receives in their first year with us is: (Required to answer. Single line text.) Enter your answer.

7. Let's think again about that typical assignment a fresh out-of-college hire receives in their first year in office. How long do they have to complete it? (Required to answer. Single choice.)
- Less than 3 hours.
- Between 3-48 hours.
- More than 48 hours.

8. Most new employees I had in mind are based in: (Required to answer. Single choice.)
- The UK.
- The EU.
- The US.
- Other Europe.

- Latin America and the Caribbean.
- Asia.

9. Curious about the results of the study? Please fill in your email below (this is optional): (Single line text.) Enter your answer.

Section 3: College Graduate (New Employee)
10. A fresh out-of-college hire (up to 1 year on their current job) in my company is typically employed at a: (Required to answer. Single choice.)
- temporary position (e.g. internship) with a fixed-term contract.
- full-time position where the new hire enjoys an open-ended contract after their probationary term.

11. We know that fresh graduates are involved in performing a variety of tasks, assignments, and projects. Some of them are routine, while others are creative. Some take just a few hours to complete, while others may last months. However, all new hires in your organisation are involved in some types of assignments more often than others. Let's call those typical assignments: the ones new hires receive most often.
The typical assignments you have received as a new employee were: (Required to answer. Single choice.)
- Routine: related to reproducing basic knowledge about the job or the field their degree is in.
- Applied: related to application of basic or more advanced knowledge in novel situations or scenarios.
- Creative: related to a critical evaluation of existing practices; suggesting novel processes, practices, or products; leading project teams.

12. The approximate share of routine tasks a new hire receives in their first year with us is: (Required to answer. Single line text.) Enter your answer.

13. The approximate share of applied tasks a new hire receives in their first year with us is: (Required to answer. Single line text.) Enter your answer.

14. The approximate share of creative tasks a new hire receives in their first year with us is: (Required to answer. Single line text.) Enter your answer.

15. Let's think again about that typical assignment a fresh out-of-college hire receives in their first year in office. How long do they have to complete it? (Required to answer. Single choice.)
- Less than 3 hours.
- Between 3-48 hours.
- More than 48 hours.

16. Most new employees I had in mind are based in: (Required to answer. Single choice.)
- The UK.
- The EU.
- The US.
- Other Europe.
- Latin America and the Caribbean.
- Asia.

17. Curious about the results of the study? Please fill in your email below (this is optional): (Single line text.) Enter your answer.

Section 4: Academic survey
18. What is the level of your course (module): (Required to answer. Single choice.)
- introductory undergraduate
- intermediate undergraduate
- advanced undergraduate
- graduate (MA/MSc, PhD)

19. We know that a final exam (or an alternative final assessment) tests a variety of skills. Some of them are fairly routine, while others are creative. Some take just a minute or two to complete, while others may last hours or days. But there is a certain type of questions on your exam that your students see most often. Let's call those typical questions.
The typical question on my final exam (final assignment) is: (Required to answer. Single choice.)
- Routine: It tests the ability of my students to understand the material (know the main facts, theories, discipline-specific formulas) that they have been exposed to numerous times during the term.
- Applied: It tests the ability of my students to apply the material to known or novel situations or scenarios, including contemporary events.
- Creative: It tests the ability of my students to critically evaluate the discipline knowledge using existing or novel techniques and, on occasion, create new knowledge.

20. The approximate share of routine questions on my final exam/assessment is: (Required to answer. Single line text.) Enter your answer.

21. The approximate share of applied questions on my final exam/assessment is: (Required to answer. Single line text.) Enter your answer.

22. The approximate share of creative questions on my final exam/assessment is: (Required to answer. Single line text.) Enter your answer.

23. Let's think again about your final exam or alternative final assessment on your course (module). How long do your students have to complete it? (Required to answer. Single choice.)
- Less than 3 hours.
- Between 3-48 hours.
- More than 48 hours.

24. Most of the students I had in mind are based in: (Required to answer. Single choice.)
- The UK.
- The EU.
- The US.
- Other Europe.
- Latin America and the Caribbean.
- Asia.

25. Curious about the results of the study? Please fill in your email below (this is optional): (Single line text.) Enter your answer.

A3. Table A-1. Within-Type Simple and Conditional Authenticity

Creative	.014	-.087
	(.081)	(.089)
Applied	-.060	-.032
	(.039)	(.047)
Routine	-.117**	-.118**
	(.048)	(.050)
Time	All	All
Location	No	Yes
Authenticity Index	-.054	-.079
Notes: Results are from regressing the share of Routine, Applied or Creative questions on indicators that respondents are employers or economics graduates. Location dummies for the UK, the USA, the EU, other Europe, and Asia are included in the second set of regressions. The dependent variable is the share of Routine questions / assignments in cells 1 and 2; the share of Applied questions / assignments in cells 3 and 4; and the share of Creative questions / assignments in cells 5 and 6. Robust standard errors are presented in parentheses. Symbols: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.		