Assessment Security and the Uncopyable Task

Designing Work That Has to Be Done in the Room

Phillip Dawson & Louise Wright

Phillip Dawson, Centre for Research in Assessment and Digital Learning, Deakin University, Melbourne, Australia

Louise Wright, Centre for Research in Assessment and Digital Learning, Deakin University, Melbourne, Australia

Assessment security is the discipline that higher education has practised without naming, and the arrival of generative models has made it urgent and visible. This article examines the design of the task rather than the policing of the student: the four properties that make a task secure, the difference between security and difficulty, the authentication of authorship through the process, the cost of the in-person component, the surveillance that institutions are tempted to buy, the equity of a design that requires a quiet room, and the indicators that a programme can collect about the reliability of its own evidence. It argues that security is a property of the design and not of the software, and that a task which can be completed elsewhere was never a task of the competence it claimed.

Contents **Page**

1. The Discipline Without a Name **24**
2. Four Properties of a Secure Task **24**
3. Security Is Not Difficulty **25**
4. The Authentication of Authorship **25**
5. The In-Person Component **25**
6. The Surveillance That Is Offered **26**
7. The Equity of a Design **26**
8. The Cost and the Hours **27**
9. The Evidence of Reliability **27**
10. The Student Who Is Honest **28**
11. The Discipline and the Task **29**
12. The Archive of the Work **29**
13. The Appeal and the Standard **29**
14. The Communication to Students **30**
15. The Sequence of a Programme **30**
16. The Reputation of the Award **31**
17. The Teacher Who Designs **31**
17. The Calendar of the Secure Task **32**

19. Conclusions

1. The Discipline Without a Name

Every programme has an implicit theory of the security of its assessments, and almost none of it is written down.

The theory consists of assumptions: that a student will not obtain help, that the work submitted was produced by the person who submitted it, that the conditions of production were comparable for the cohort. The assumptions have held because the cost of breaking them was high, and the cost has fallen to the price of a subscription.

The consequence is that the security of an award now depends on an assumption that nobody has examined and that the institution has not tested. The instrument is to write the theory down: for each task, what it assumes about the conditions of production, and what evidence the programme would produce if the assumption were challenged.

2. Four Properties of a Secure Task

A task is secure when four properties hold, and a task that lacks one of them can be completed by somebody other than the student.

The first is observation: the work is produced under conditions that the institution controls. The second is authentication: the programme can connect the work to the person, through a defence, a version history or a recorded process. The third is specificity: the task depends on material that is particular to the course — the data set of the laboratory, the case of the seminar, the placement of the student. The fourth is unpredictability: the task is not the same for every candidate, so that a solution cannot circulate.

The four can be combined and each of them raises the cost of a substitution. The instrument is a template of the assessment of a programme in which each task is scored on the four properties, and the score is the map of the programme's exposure.

3. Security Is Not Difficulty

The most common error in response to a new threat is to make the task harder, and difficulty is not security.

A harder essay is still an essay that can be produced elsewhere. A harder examination question is still a question whose answer can be memorised. The increase in difficulty raises the proportion of students who fail and it does not raise the proportion whose work is their own, and the two effects are unrelated.

The instrument is to separate the two questions in the design: what competence the task certifies, and what evidence connects the work to the student. A task can be easy and secure, which is the laboratory exercise done in a session; it can be hard and insecure, which is the dissertation written at home with no defence.

4. The Authentication of Authorship

The instrument with the strongest evidence behind it is the authentication of authorship through the process, and the process is what a programme usually discards.

A portfolio of three drafts with the comments of a teacher, a logbook of a laboratory, a version history of a document, a recorded presentation of a project in progress: each of them produces a trace that a substituted text cannot supply. The trace is also the most useful material for teaching, because it shows the student where her argument changed.

The consequence is that the instrument which secures an assessment is the same instrument that improves it. The programme that requires the process obtains the evidence of authorship and the material of feedback in a single act, and the cost is the design of the stages.

5. The In-Person Component

An in-person component is the most reliable security instrument available and the one with the largest logistical cost.

A defence of twenty minutes, an oral examination of a group, a practical in a laboratory, a presentation with questions: each of them requires a room, a member of staff and a calendar, and each of them pro-

duces evidence that cannot be fabricated elsewhere. The cost per student is a multiple of the cost of marking an essay, and the multiple is the reason that the component was abandoned in the first place.

The instrument is to use the in-person component where it certifies the most: the capstone of a programme, the professional practice of a clinical or engineering course, the final assessment of a language. A programme that defends its capstone and not its first-year essay has allocated the expensive instrument where it matters.

6. The Surveillance That Is Offered

The market has offered institutions a set of instruments of surveillance — the proctoring software that watches a student through a camera, the browser that locks her machine, the detector that scores a text — and each of them carries a cost that is not in the price.

The proctoring software requires a room of her own, a device that meets a specification and a connection that does not fail, and it excludes the student who has none of the three. It records the face of a student in her home and transmits it to a company. The detector produces a probability that has been shown to be unreliable and that falls unequally on students writing in a second language.

The instrument is a rule that a surveillance technology may be adopted only if it is shown to be accurate, proportionate and equally available to every student, and that the burden of the demonstration rests on the institution that wishes to buy it. No instrument offered so far has passed the test.

7. The Equity of a Design

A secure design distributes its costs unequally unless it is designed not to, and the inequality falls on the students with the least room of their own.

A portfolio with three drafts requires a quiet place to write them. An in-person examination requires the ability to attend at a fixed hour, which a working student may not have. A proctored examination at home requires a room and a device. Each of the four properties of a secure task has a cost that a student with a household may not be able to bear.

The instrument is an equity review of every redesigned assessment: which students cannot meet the conditions, and what alternative the programme offers that certifies the same competence. The review is a page in the file of the programme and it is the condition under which security does not become a filter.

8. The Cost and the Hours

The redesign of an assessment is measured in hours of staff time, and the hours are the binding constraint in every programme that has attempted it.

A task that requires a defence of twenty minutes for a cohort of two hundred requires sixty-seven hours of staff time, which is the equivalent of a course. A portfolio with three stages requires the marking of three submissions and the comments at each stage, which is three times the marking of the final version. The security is real and it is not free.

The instrument is to fund the redesign as a project with a stated number of hours per course, allocated to the teachers who do it, and to accept that the number of assessments of the programme will fall. A programme that secures every task and keeps the same number of them has not made a decision; it has added a cost.

Property	What it prevents	Instrument
Observation	Work done elsewhere	A session under supervision
Authentication	A substituted text	A portfolio with drafts
Specificity	A generic solution	Material particular to the course
Unpredictability	A circulating answer	A task generated per candidate

Table 1 Four properties of a secure task and the instrument of each

9. The Evidence of Reliability

A programme can measure the reliability of its own assessment evidence, and the measurement is a sample and a comparison rather than a survey.

The instrument is a reliability audit: a sample of the work of the graduates, marked independently by two readers, with the agreement between them recorded, and a comparison of the sample with the

standard of the discipline. A programme that has done the audit knows whether its marks mean the same thing in two classrooms, which is the question that the students and the employers are asking.

The audit is also the instrument through which a programme detects the task whose marks are inflated, which is the task that requires the most attention and that no questionnaire identifies.

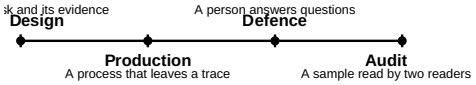

Figure 1 The sequence of a secured assessment

10. The Student Who Is Honest

The design of security affects the honest student more than the dishonest one, and the effect is the one that the institution never measures.

An honest student who is required to submit three drafts, attend a defence and accept a proctored examination has paid a price in time and in anxiety for a problem that she did not create. A student who is willing to buy an essay has paid a smaller price and has frequently obtained a better mark, because the essay she bought was written without the constraints of her week.

The consequence is that an assessment system which is insecure rewards the student who breaks the rule and penalises the one who keeps it. The instrument is to make the secure task the ordinary task, so that the honest student is not paying a surcharge for the integrity of the system.

11. The Discipline and the Task

The form of a secure task is different in each discipline, and the difference is the reason that a central policy cannot design it.

In engineering, the secure task is a design produced in a studio with the drawings reviewed. In law, it is a memorandum written in a session with the statutes available. In medicine, it is a clinical examination with a patient and an examiner. In history, it is a defence of an argument against the questions of two readers who have read the same sources.

The instrument is a disciplinary catalogue of secure forms, maintained by the academic community of each field, with an example of each form and a description of what it certifies. A catalogue of that kind is the instrument through which a programme finds a form that fits its competence instead of inventing one in a crisis.

12. The Archive of the Work

A programme that keeps the work of its students for a period has an archive, and the archive is the instrument through which a challenge can be answered.

A case of suspected substitution is decided on evidence, and the evidence is frequently a comparison with the previous work of the same student. An archive that holds the work of two years permits the comparison; an archive that is deleted at the end of each semester does not.

The instrument is a retention rule: the assessments of a programme are kept for the duration of the appeal period and for one further year, in a form that permits the comparison and respects the rights of the student. The rule is administrative and it is the condition under which a procedure can be fair to both parties.

13. The Appeal and the Standard

An appeal against a finding of misconduct tests the standard of evidence that the programme actually applies, and the standard is frequently lower than the regulation states.

The regulation says that a finding must rest on evidence that would satisfy a reasonable person. The practice is that a marker's suspicion

and a detector score are treated as sufficient, and the appeal body has neither the time nor the expertise to examine the text. The consequence is a system in which the outcome depends on whether the student appeals.

The instrument is a standing panel with a technical member, a published statement of the evidence that a finding requires, and a rule that the panel must state what it relied on. A panel of that kind produces a small number of well-founded findings instead of a large number of contested ones.

14. The Communication to Students

The security of an assessment depends on the students understanding what is required and why, and the communication is usually a paragraph of a regulation.

A student who understands that a defence exists to certify her own work, and not to catch her, prepares for it as a professional exercise. A student who receives a rule without a reason treats it as an obstacle to be managed, and the management is the beginning of the behaviour the rule was meant to prevent.

The instrument is to teach the purpose in the first session of the course: what the task certifies, what evidence the programme will rely on, and why the secure form serves the student rather than the institution. Two slides in a first lecture are cheaper than a detector and more effective.

15. The Sequence of a Programme

A programme that decides to secure its assessment can do so in a sequence, and the sequence begins with the map rather than with a purchase.

The first step is the map: each task scored on observation, authentication, specificity and unpredictability. The second is the redesign of the two or three tasks with the weakest evidence, funded in hours. The third is the in-person component at the capstone, with the rooms and the calendar arranged a year in advance. The fourth is the reliability audit and the archive.

The fifth is the rule on surveillance technologies, adopted before a vendor makes the decision for the institution. A sequence of that kind takes two years and it produces a programme that can defend its own award.

16. The Reputation of the Award

The security of assessment is a question of the value of the degree, and the value is a collective good that no single institution can protect alone.

An employer who hires a graduate of a programme that cannot certify its own work learns to discount the degree, and discounts the degrees of the system with it. The consequence is that an institution which reduces its standards to attract students damages the asset that attracts them, and the damage is invisible for a decade.

The instrument is a sector-wide statement of the minimum evidence that an award requires, adopted by the conference of rectors and published, with the institutions reporting their compliance. A statement of that kind is the only defence of a collective good against an individual incentive.

17. The Teacher Who Designs

The competence that the reform requires is the design of an assessment, and it is the least taught competence of the academic profession.

A teacher who has never designed a task that certifies what it claims will reproduce the tasks that she experienced as a student, and those tasks were designed for a world in which the substitution was expensive. The reproduction is not a failure of intelligence; it is the absence of a formation.

The instrument is to make the design of assessment a part of the preparation of every new member of staff, taught with the material of her own discipline, and to recognise it in promotion. A programme that has a teacher who can design is a programme that can respond to the next change; one that has a policy is a programme that will be surprised again.

17. The Calendar of the Secure Task

A secure assessment is a calendar before it is a design, and the calendar is what most programmes discover too late.

A defence of twenty minutes for a cohort of two hundred requires a fortnight of slots, two rooms and four members of staff; a portfolio with three stages requires three deadlines that do not collide with the assessment of the other modules; a supervised session requires a hall, an invigilator and a period in which no other examination is scheduled. Each of the four is a constraint on the calendar of the institution.

The instrument is to design the calendar of the programme before the tasks: which weeks contain the in-person components, who is available in each, and what the students are asked to produce between them. A programme that has the calendar can adopt the secure forms; one that does not will adopt them for the first cohort and abandon them for the second.

The calendar is also the instrument through which the cost becomes visible. An institution that sees the number of staff hours that its secure assessments require has the information on which to decide how many of them it can afford, and the decision is honest when it is taken in the calendar rather than in the crisis of the marking period.

19. Conclusions

Assessment security is a property of the design of a task and not of the software that polices it, and a task that can be completed elsewhere has certified nothing about the person who submitted it.

Three conclusions follow. The map comes first: each task scored on observation, authentication, specificity and unpredictability, and the two or three with the weakest evidence redesigned and funded in hours. The in-person component comes second, at the capstone and in the professional courses, where it certifies the most and costs the most. And the evidence comes third: a reliability audit of a sample of the graduates' work, an archive that permits a comparison, an appeal standard published, and a rule that no surveillance technology is adopted without a demonstration of its accuracy and its equity.

An institution that has done the three can defend its award. An institution that has bought a detector has bought a suspicion and called it a standard.

References

[1] Dawson, P. (2021). Defending assessment security in a digital world. London: Routledge.

[2] Dawson, P., & Bearman, M. (2023). Assessment security in the age of generative AI. Assessment and Evaluation in Higher Education, 49(1).

[3] Bretag, T. (2016). Handbook of academic integrity. Singapore: Springer.

[4] Joint Information Systems Committee (2023). Generative AI and assessment security. Bristol.

[5] Quality Assurance Agency (2023). Assessment in the context of generative AI. Gloucester.

Biography:

Phillip Dawson is a professor at Deakin University working on assessment security, academic integrity and the design of assessment in higher education.

Louise Wright is a researcher at Deakin University working on assessment design, authorship and the evidence of student learning.