

Developing a Code of Practice for Using Data in Wellbeing Support

Andrew Cormack¹, David Reeve²

Abstract

With student and staff wellbeing a growing concern, several authors have asked whether existing data might help institutions provide better support. By analogy with the established field of Learning Analytics, this might involve identifying causes of stress, improving access to information for those who need it, suggesting options, providing rapid feedback, even early warning of problems. But just investigating the possibility of such uses can create significant risks for individuals: feelings of creepiness or surveillance making wellbeing worse, inappropriate data visibility destroying trust, assessments or interventions becoming self-fulfilling prophecies. To help institutions decide whether and how to explore this area, and to reassure individuals that this is being done safely, we propose a Wellbeing Analytics Code of Practice. This starts from an existing Learning Analytics Code, confirms that its concerns and mitigations remain relevant, and adds additional safeguards and tools for the wellbeing context. These are derived from a detailed analysis of European and UK data protection law, extracting all rules and safeguards mentioned in relation to health data. We also develop context-specific tools for managing risk and evaluating data sources. Early feedback suggests that these documents will indeed increase confidence that this important area can be safely explored.

Notes for Research

- Prior work identified ethical and legal concerns when using data to enhance learning processes (“learning analytics” and “curriculum analytics”) and developed a Learning Analytics Code of Practice (Sclater & Bailey, 2018) to address these. More recently the need for, and possibility of, “wellbeing analytics” using data to enhance student and staff wellbeing processes has also been identified.
- This paper describes how an in-depth study of European data protection law was used to develop a complementary Wellbeing Analytics Code of Practice (Cormack & Reeve, 2020), building on the Learning Analytics Code, to guide the investigation and routine use of data by institutions to improve support for wellbeing.
- We conclude that it should be possible to research, investigate, and conduct wellbeing analytics ethically, lawfully, and safely. First, this must be led by health theory (as learning analytics should be led by pedagogical theory). Second, design and testing of models and systems are critical and best viewed as independent processes. Third, a wider range of data sources will likely be required, but this requires particular care. Although our work is based on European law, that law is increasingly considered a global “gold standard,” so we believe our work should be applicable elsewhere.
- These should be fruitful topics for research, and important enablers for practice, with the Code of Practice providing a framework within which both research and practice can be conducted safely.

Keywords

Learning analytics, wellbeing, data protection

Submitted: 06/07/21 — **Accepted:** 05/01/22 — **Published:** 31/05/22

Corresponding author ¹ Email: andrew.cormack@jisc.ac.uk Address: Jisc, Lumen House, Library Avenue, Harwell Oxford, Didcot OX11 0SG, United Kingdom. ORCID ID: <https://orcid.org/0000-0002-8448-2881>

² Email: david.reeve@jisc.ac.uk Address: Jisc, 4 Portwall Lane, Bristol, BS1 6NB, United Kingdom. ORCID ID: <https://orcid.org/0000-0001-9814-6734>

1. Context

Even before the pandemic’s additional impact (Office for National Statistics, 2021), the wellbeing and mental health of university and college staff and students was a growing concern. The Higher Education Academy’s Student Academic Experience Surveys in 2016 and 2017 (Neves & Hillman, 2017) found students reporting wellbeing measures 10% to 20%

below the equivalent age group across the UK population and that this gap was increasing. A National Union of Students survey for the All-Party Parliamentary Group on Students (Gil, 2015) found that 78% had experienced mental health issues in 2015. Although part of the decline may be a wider societal trend affecting this age group (Education Policy Institute, 2018), students may be particularly at risk and, through their close relationship, educational institutions may be particularly able to help (Thorley, 2017).

Many universities and colleges already use data to inform, enhance, and accelerate their support for tutorial difficulties. Noting that some tutorial problems revealed wider wellbeing issues, authors such as Ahern (2018) and Hall (2019) suggested that data might be incorporated directly into wellbeing support processes. There might be opportunities to enhance existing on-demand and event-based support (e.g., University of South Wales, n.d.) with wellbeing equivalents of tutorial practices such as avoiding data re-entry (Universities UK, 2021), helping students compare their experience with others (Hughes & Spanner, 2019), informing students and staff about progress (Verbert et al., 2020), personalized triage suggesting appropriate resources and approaches — whether self-help, within the institution, or from external services (Abdi et al., 2020) — even early warning of signs of problems (Lang et al., 2020). Alternatively, in the same way that data from individuals can be used to identify problems with materials and courses (Kaliisa et al., 2020, p. 41), data might highlight systemic causes of poor wellbeing. These could then be addressed, for example, by reducing peaks in cross-course workload (Perkins et al., 2020), or mitigated by offering support at pinch points such as first mark, January gloom, or re-sits (Jones, 2019; University of South Wales, n.d.). As the field of data-informed tutorial support — extending well beyond statistics to include information, triage, and alerting — had been christened “learning analytics” (Corrin et al., 2019), so the broad potential for data-informed wellbeing support was referred to as “wellbeing analytics.”

Comparing the two concepts, wellbeing support clearly involves both greater potential benefits and greater risks. Early and effective resolution of wellbeing problems could prevent them developing into significant mental and physical health issues: even, in a few tragic cases, self-harm and suicide. But the information needed to support such beneficial activities is more sensitive, possibly including existing and developing health issues, activity and sleep patterns, and engagement with counselling and health services. Inaccurate inferences about an individual’s state of mind could even become a “pathogenic” self-fulfilling prophecy (Prinsloo & Slade, 2016; UHI, 2017).

Another difference is the range of data sources likely to be needed. Learning analytics is typically a closed system: data gathered during learning processes helps to improve those same processes. This should be a relatively obvious, perhaps even expected, reuse of such data. To support wellbeing, however, there is no equivalent primary data source. We might plausibly need both “environmental” information about the stresses individual students experience and “behavioural” information about how they respond to those stresses. Environmental stresses are likely to include some, such as workload, created by the institution. But accommodation and financial issues (Jisc, 2019) as well as social and family circumstances (Mulenda, 2020) are often cited as important for wellbeing. Behavioural signals might appear in academic work, but attendance, use (or non-use) of spaces (Ahern, 2018), engagement with student societies (Hughes & Spanner, 2019), quantity and timing of activity (Ahern, 2017), even mobile phone use, sedentary time, and sleep patterns (Mack et al., 2021) may be earlier and more accurate indicators of difficulty. These are much more intrusive than learning data to request or gather, much more sensitive to hold, and can cause much more harm if misused. Although their relevance may be clear in retrospect, students and staff may find it instinctively surprising, even shocking, that their institution wants to know. Even where institutions already have relevant information, using it for new purposes can “substantially” change the ethical issues raised (Corrin et al., 2019, p. 6).

Even exploring which data sources might be informative and what they can add to existing approaches — in effect, investigating whether “wellbeing analytics” is worth pursuing — raises significant ethical and practical issues. Both Slade and Prinsloo (2014, p. 299) and Ferguson (2019, p. 25) suggest that we may feel “a moral responsibility for employing information which aims to provide more effective and relevant [tutorial] support for all students.” But Selwyn (2019) fears that analytics may be perceived as “dataveillance” of students and staff: “a means of surveillance rather than support,” resulting in a “pushback against learning analytics” and other secondary uses of student data. Other authors highlight broader risks in using data: that historic data may reinforce historic biases and discrimination (D’Ignazio & Klein, 2020, chapter 6); that systems informed by current data may disadvantage those who — for personal, social, behavioural, or technical reasons — have a smaller data footprint (Sclater, 2016, p. 35); and that future over-reliance — whether by institutions or individuals — on data-informed decisions may reduce personal autonomy (Prinsloo & Slade, 2016, p. 162). More positively Prinsloo and Slade (2016, p. 177) suggest that engaging students as participants in “collaborative processes of sense-making” around data would increase agency. As well as these risks of data misinterpretation and misuse, we must remember that simply collecting data — if perceived as intrusion or surveillance — could impact wellbeing or result in counterproductive behaviour changes by students and staff (Ahern, 2020, p. 8). Using data to support wellbeing must not result in individuals concealing their struggles, or put them off seeking advice or help when it is needed. Knowing when NOT to use data is at least as important.

Such concerns are not limited to unauthorized data reuse by learning system vendors (Corrin et al., 2019, p. 4) or supposedly consented to apps (McKie, 2019) and may arise even where there is no personal data collection or intervention (e.g., Kitchener,

2016). More fundamentally, Slade, Prinsloo, and Khalil (2019, p. 242) worry that, without transparency and good governance, institutional data reuse could undermine students’ “inherent trust in their university to use their data appropriately and ethically” (p. 235). Wachter and Mittelstadt’s “Right to Reasonable Inference” (2019, p. 581) may help guide and explain choices of data source, but institutions must be open to those choices being challenged. Indeed Griffiths and colleagues (2016, p. 23) suggest, “The protection of privacy should not be seen as a burden but rather as a valuable service that can be offered to build trusting relations with our stakeholders.” Losing stakeholder trust can have reputational and operational consequences if it means that institutions cannot obtain or process the data needed for core functions. For example, loss of trust in InBloom’s data reuse practices left many US schools unable to use their primary administrative system (Griffiths et al., 2016, p. 3).

Trustworthy systems must be based on institutional decisions, not on leaving individual students to assess the complex implications of using or not using data (Corrin et al., 2019, p. 4; Prinsloo & Slade, 2016). The process of making those decisions and designing those systems can itself build confidence if relevant communities and individuals are involved, in particular in the following aspects:

- Deciding which, if any, applications of data-supported wellbeing fall within the institution’s role, remit, culture and capabilities
- Thinking carefully about processes and safeguards to ensure they are sound from a medical, privacy and, where appropriate, statistical perspective
- Ensuring the risks to individuals from data use are understood and minimized and that the potential benefits justify the remaining risks
- Monitoring processing and outcomes to ensure that the initial assessment of risk and benefit was correct
- Explaining these processes clearly to all individuals and communities they affect

Every institution will reach its own conclusions on whether and how to use data to enhance its existing wellbeing support. Some might use timetable data to identify and mitigate high-stress periods for a whole cohort (Perkins et al., 2020). Some might improve data accessibility so students do not have to explain their situation when they visit support services (Universities UK, 2021). Some might consider offering personalized self-help, alerts, or interventions (Inkster et al., 2021). As the preceding discussion indicates, this choice will depend on many factors, including the following:

- The existing wellbeing provisions and resources
- The estimated unmet need and capacity to satisfy them
- The availability of good-quality data
- The organizational and technical structures within which increased, but controlled, data flows could occur
- The institution’s current uses of data to support other processes
- The attitudes of students, staff, and departments to those uses
- The different experiences and expectations of health privacy among the student community
- The level of data literacy among students, staff, developers, and management

An institution that already has a standardized approach and governance for using student activity data, students and staff who expect this to be used to provide rapid feedback, and a well-resourced internal wellbeing service should reach very different conclusions from a department that merely has timetable data.

However, in exploring these issues, all institutions are likely to consider similar issues including governance and decision-making, capability (including data quality and availability), assignment of roles, risk assessment and balancing, transparency, and communications. We therefore saw an opportunity for a common document that could guide those discussions, decisions, and any subsequent implementations. By taking a comprehensive view of the field and documenting good practice within it, this could reassure students, staff, and management that appropriate choices are being made.

2. Methods

In a similar situation in 2015 — where institutions wanted to use data to improve learning, but were concerned whether and how this could be done safely — Jisc worked with universities and colleges to develop a Learning Analytics Code of Practice (Sclater & Bailey, 2018). This fitted Jisc’s purpose and relationships as a non-profit organization helping UK tertiary education use technology safely and effectively (Jisc, n.d.). Sclater (2016) conducted an extensive review of academic literature that identified ninety-three ethical or legal concerns. A panel of experts from student union representative to pro-vice-chancellor then grouped those concerns — from stewardship and validity to enabling positive interventions and minimizing adverse impacts — and developed good practice guidance for how they might be addressed. The resulting Code of Practice has been used by institutions to explore whether and how to adopt learning analytics approaches, and to ensure that they are subject to appropriate safeguards in design, operation, and review. Indeed the National Union of Students made it their key reference in advising student unions on how to approach learning analytics issues (NUS, 2015).

We expected that work to apply equally to wellbeing so, rather than duplicate it, we added a source not considered by Sclater: statute law, in the form of Europe’s General Data Protection Regulation (henceforth GDPR). We already knew this offered guidance and processes for two new wellbeing concerns: risk assessment and diverse data sources. It also has specific provisions and safeguards for using data about health and for health purposes. Our GDPR review therefore aimed 1) to confirm that Sclater’s work applied to this new context and 2) to identify additional tests and safeguards to consider for data-supported wellbeing.

The GDPR is well suited to this use, having the dual aims of protecting natural persons while facilitating the beneficial use of personal data, and an approach focused on processes and safeguards rather than prescriptions. Concepts such as Data Protection by Design and Accountability match our code’s aim of encapsulating responsible behaviour. Furthermore, GDPR directly addresses the new challenges we identified in moving from learning to wellbeing: high stakes risk and benefit through Data Protection Impact Assessments, and diverse data sources through Purpose Compatibility. Although it only formally applies in Europe, the GDPR’s approach is being adopted in other regions and countries (Woodward, 2021). We hope that our work will have similar wide relevance.

Note that we were not aiming to identify a single path through the GDPR that ensures “compliance.” Even if that were possible — and the range of activities and contexts comprising data-supported wellbeing makes that unlikely — we wanted more from the legislation. Extracting all the conditions and safeguards mentioned for health data or processing lets us create a single framework for many wellbeing developments, such as these:

- Which kinds of data-enhanced wellbeing support should an institution consider?
- Which will be too risky or onerous?
- Which data sources should be investigated?
- What resources might be needed?
- What safeguards should be applied?
- Which roles need to be considered?
- How can decision-making and continuing development be governed?

Including all relevant safeguards and good practices, rather than just those required by law, might seem overcautious. But in an area as sensitive as wellbeing we believe an approach that explicitly goes beyond the legal requirements to demonstrate safety and effectiveness is a better foundation for trust. Once an institution has chosen a particular approach, it can revisit the law as a compliance tool for that specific activity.

We therefore considered how all 99 Articles and 173 Recitals of the GDPR might apply to institutions conducting data-supported wellbeing activities. Where that law invites more detailed member state legislation, we reviewed the UK’s additions (as the state with which we are most familiar) in the Data Protection Act 2018 (henceforth DPA18). Particularly relevant were the Articles and Recitals containing GDPR’s seven principles (Article 5); its six lawful bases for processing personal data (Article 6); and the rights it grants to individual data subjects (Chapter III, Articles 12 to 23). As well as these general provisions, we learned from the ten lawful bases (Article 9) for “Special Category” processing (formerly known as “Sensitive Data” and including health). These describe situations where such processing may be lawful, and the conditions and safeguards that must be applied. Emergency provisions — such as those on life-saving treatment (Art. 9(2)(c)) and pandemics (Art. 9(2)(i)) — were less informative as these reduce, rather than add, safeguards. Further detail came from GDPR Article 89 on research and statistics and the UK’s more detailed specifications in Section 19 and Schedule 1 of the DPA18. Relevant guidance documents from the UK (henceforth “Information Commissioner”) and the EU (henceforth “Article 29 Working Party”) data protection regulators were also considered.

Our developing ideas were shared through blogs and mailing lists, and discussed in meetings with the Office for Students, data protection officers, learning analytics practitioners, and the Association of Student Support Services (AMOSSHE). All were positive and provided useful feedback.

3. Results

Sclater and Bailey (2018) grouped their legal and ethical concerns into eight headings: 1) responsibility, 2) transparency and consent, 3) privacy, 4) validity, 5) access, 6) enabling positive interventions, 7) minimizing adverse impacts, and 8) stewardship of data. Reviewing the legal requirements and expectations for Wellbeing Analytics, we concluded two things: first, that the concerns addressed in their Learning Analytics Code were still relevant; second, that those headings still provided an appropriate structure.

The following sections explain how our approach confirms the relevance of Sclater’s work to wellbeing, the additional detail and safeguards we identify, and how those are incorporated into our Wellbeing Code (Cormack & Reeve, 2020). There are important new tools: Purpose Compatibility supporting Transparency and Consent (see 3.2.1 below and Cormack & Reeve, 2020, Annex B) and Data Protection Impact Assessments supporting Enabling Positive Interventions/Minimizing Adverse

Impacts (see 3.6.1 below and Cormack & Reeve, 2020, Annex A). Health data reverses the normal presumption of Access (see 3.5 below).

3.1. Responsibility

Sclater (2016) “aims to make it clear that responsibility in the institution for the various processes of learning analytics needs to be allocated appropriately and that students and other stakeholders should be consulted” (p. 32). Clear institutional responsibilities are key to the GDPR’s Accountability Principle (Art. 5(2)); consultation should form part of Data Protection Impact Assessments (DPIAs; see 3.6.1 below). This aim, and section of the Learning Analytics Code, are therefore equally valid for wellbeing.

UK law is specific that before processing to “protect [...] the physical, mental or emotional wellbeing of an individual” the institution must identify an unmet “care and support” need that the processing will assist and that the institution has (or can obtain) the resources to address. Institutions should also consider whether the need is more appropriately met by another organization (see 3.7 below). The identified risk may affect either a particular individual or some among a “type of individual” such as students. Processing data about those in a cohort who do not currently need help may be permitted if it is necessary to identify and provide better treatment to those who do (Article 29 Working Party, 2007, p. 5), though strong safeguards must be applied (DPA18, Sch. 1 Part 2 Para. 18).

This might require processing data without individual consent, for example to help a cohort, or to validate risk assessment methods (GDPR Recital 54 mentions, “determinants having an effect on health status”), or if we know that those in serious need are not coming forward. However explicit individual consent must be sought as soon as that becomes an appropriate basis: typically, when an individual is identified as being at risk (Art. 9(2)(a) and 3.2 below). The reasons for starting without consent must be documented and justified.

Whereas the use of data to support teaching should be led by teaching professionals (Ferguson et al., 2019), the GDPR’s provision for “preventive or occupational medicine” (Art. 9(2)(h)) indicates that the choice, design, provision and review of data-supported wellbeing activities should be done “by or under the responsibility of” health professionals. Other staff — such as tutors — are likely to be involved in identifying student needs, but must have appropriate training, skills and resources to perform this role confidently and safely, working under supervision and subject to confidentiality rules. Given the range of different interactions between institution and student — teaching/learning, accommodation and social life, as well as providing wellbeing support — these rules and processes must make clear that wellbeing data is only used for wellbeing purposes (Article 29 Working Party, 2007, p. 11). Any new wellbeing role assigned to tutors, or other staff, must not be perceived as conflicting with their existing roles and relationships.

Finally, the GDPR Accountability principle (Art. 5(2)) requires not only that institutions follow clear processes — satisfying the other principles when designing, documenting, and monitoring processing, but that they are able to prove they have done so. This has particular benefits in sensitive areas, like wellbeing, where trust is essential and easily lost. Three GDPR documents help demonstrate trustworthiness, and may be legal requirements for health processing:

1. During design, Data Protection Impact Assessments involve considering and consulting on the likely risks and benefits to individuals (Art. 6(1)(f), Art. 35/36; Information Commissioner’s Office, 2019a; Information Commissioner’s Office, 2019b)
2. During operations, Records of Processing Activity ensure that designed safeguards are working (Art. 30; Information Commissioner’s Office, n.d.)
3. During monitoring, review, and information retention/disposal processes, an Appropriate Policy Document ensures that proper procedures are followed (DPA18 Schedule 1 Part 4; Information Commissioner’s Office, 2019e)

3.2. Transparency and Consent

Under this heading, Sclater includes both “being open about all aspects of the use of learning analytics, and ensuring that students provide meaningful, informed consent” (2016, p. 32). Transparency is a GDPR Principle (Art. 5(1)(a)), with specific requirements on transparency processes in Articles 13 and 14. Lists of information that should be provided are in those two Articles and Article 6(1)(e), Public Interest lawful basis.

Consent is usually the appropriate lawful basis for self-reported data and individual interventions, but has significant disadvantages, both for the institution and the individual, if used too early in a comprehensive system. Even for health data, the GDPR notes situations where consent is inappropriate, including “the provision of, and universal access to, health care” (Rec. 54) and “quality control, management information and [...] local supervision” of care systems (Rec. 53). Sclater (2016, p. 32) points out that “[a]llowing students to opt out of data collection may, in some cases, make the carrying out of normal educational processes impossible.” Valid consent requires “an affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject’s agreement” (Rec. 32). For health data, consent should also be “explicit” (Art. 9(2)(a); Information Commissioner’s Office, 2019d), but early consent is often broad, under-informed, ambiguous and —

when sought from students by the institution they depend on (Rec. 43) — not free. Consent should, by definition, produce incomplete datasets and a high risk that models and processes will be biased or discriminatory. Rather than blanket “wellbeing support,” seeking consent for specific interventions allows clear explanation of their implications. Individuals may decline one intervention without ruling out others (Cormack, 2016). The Accountability Principle plays a key role in deciding whether, when, and how to seek consent. Even where consent is used, the safeguards set out in the rest of the Code are essential to ensure that individual choice is appropriately protected.

Consent may be withdrawn at any time. The GDPR also provides a right to object where processing for a public or legitimate interest (Art. 21). Formally this only requires a review of the risk/benefit balance for the specific individual but, where processing is supposed to benefit that individual’s wellbeing, objections should probably be treated as unconditional opt-outs. Even if the law allows an objection to be overridden, doing so seems more likely to harm than improve wellbeing. Furthermore, if an individual does object, refuse, or withdraw consent, we should review our information and processes to see why they did not feel comfortable accepting the offer of help.

As discussed by Edwards and Veale (2017) and Binns (2019), the possibility of objections does not justify retaining personal data “just in case” it is needed to rebuild statistical models.

3.2.1. Purpose Compatibility

Supporting wellbeing is likely to require a wider range of data sources than supporting learning. Sources of stress might include accommodation, finance, and family background. Individual responses to stress might appear in their use (or non-use) of campus services such as libraries, accommodation, or social and health services. Some of these will be too intrusive to reveal to tutors or support services unless the student volunteers the information; some may have less intrusive, but still effective, alternatives. While Wachter and Mittelstadt’s (2019) “plausible connection” rule is a good start, the GDPR Purpose Compatibility Principle (Art. 5(1)(b) and Art. 6(4)) provides a specific mechanism to identify the most appropriate — which may not mean the most statistically powerful — data sources to use. Institutions must consider how different the new purpose — here, individual wellbeing support — is from the purpose for which the information was originally collected. That might, for example, have been individual health support, individual tutorial support, other individual support (e.g., financial), operational, or statistical purposes. The difference between the original and new purposes provides a strong indication — reflecting Recital 50’s “reasonable expectations” test — whether or not a source should be used for wellbeing support.

Some potential data sources will already be held by the institution, either as a result of transactions (e.g., learning platforms or library loans) or observed during use of campus facilities (e.g., WIFI networks); others, such as finance, may require agreements with other organizations. These three kinds of sources present different — probably increasing — challenges for Recital 50’s requirement that individuals be informed of new uses of their data. Transactions offer an obvious opportunity to present that information; observed data may require notices, with no confirmation that the individual has seen them; commitments may be needed from third parties that they will provide the information when data are collected. Combining this “difficulty of informing” measure with purpose (in)compatibility produces a heat map highlighting data sources that are compatible/easy-to-communicate and therefore preferable to incompatible/hard-to-communicate ones (Cormack & Reeve, 2020, Annex B). Where a proposal can only be achieved with “difficult” data sources, this may indicate that it should not be taken forward.

3.3. Privacy

The GDPR’s Integrity and Confidentiality Principle (Art. 5(1)(f)) confirms Scater’s requirement “that access to student data should be carefully controlled and data protection legislation complied with” (2016, p. 33). The Substantial Public Interest lawful basis requires “suitable and specific measures to safeguard the fundamental rights and interests of the data subject” (Art. 9(2)(g)).

Recital 54 warns that “processing of data concerning health for reasons of public interest should not result in personal data being processed for other purposes by third parties such as employers or insurance and banking companies.” The Article 29 Working Party (2007) adds that identifiable health data should be visible only — unless the individual agrees — to those directly involved in treatment. The importance of this separation for trust between students and institutions is highlighted by a BBC (2019) article where students avoided university wellbeing services for fear that using them might tarnish their academic record.

Health data and assessments represent a high risk, so data security must be “appropriate[ly]” strong: for electronic health records the Working Party (2007, p. 19) considers, “Access by unauthorized persons must be virtually impossible and prevented.” Clear instructions reduce the risk of mistakes by those with authorized access; access logs and regular audits help detect mistakes and unauthorized access. Measures should be used at all levels — from technical to training and organizational — to prevent inappropriate flow, use, or visibility of wellbeing data.

3.4. Validity

Sclater (2016) discusses the validity of predictive learning analytics algorithms. GDPR Recital 71, too, draws attention to the validity of “mathematical and statistical procedures.” But in wellbeing, the serious consequences of errors mean all processes, not just numerical ones, should be tested, monitored, and reviewed.

These three activities may need more data than is required to operate a wellbeing support system. Testing should evaluate potential data sources that are found not to be sufficiently informative. Monitoring may need additional, special category data to identify signs of bias or discrimination. Review may need to consider historical data. By treating these as separate “research” processes, the GDPR provides additional safeguards to mitigate that increased risk. Such processes should be carried out using pseudonymized or fully anonymized data (Art. 89(1)). Information that they use or discover must not be used for “measures or decisions regarding any particular natural person” (DPA18, s. 19(3)). In terms of data flows, however, testing, monitoring, and review are one-way streets: data may go in, but only improved processes can come out.

Treating testing, monitoring, and review as separate purposes from the operation of wellbeing support also empowers the GDPR’s Data Minimization and Storage Minimization Principles (Art. 5(1)(c) & (e)). These permit only “necessary” storage and processing of data, in the sense that any less data or shorter retention would prevent the purpose being achieved (Article 29 Working Party, 2014). Since each purpose requires different data, collected and used over different time periods, their processes should retain different types of data — whether anonymous, pseudonymous, or identified — for different lengths of time.

Finally, considering — and, if possible, conducting — these processes separately should reduce the risk of data leakage between them. Testers and reviewers should not be able to see individual identities: those providing support, who do need to know identities, should not have access to information required only for the testing, monitoring, or review stages.

3.5. Access

Access is the one significant change between learning analytics and wellbeing. For health data, UK law reverses the GDPR Article 15 presumption that individuals can obtain their own data. Health data may only be disclosed once a health professional has confirmed that this is not “likely to cause serious harm to the physical or mental health of the data subject or another individual” (DPA18, Schedule 3 Part 2 Paragraph 5).

3.6. Enabling Positive Interventions/Minimizing Adverse Impacts

For learning analytics, Slater (2016) highlights the challenge of balancing any “obligation to act” with the “potential pitfalls of intervening directly with students based on analytics” (p. 34). There are “potential detrimental effects” if processes or data “categoris[e]” students, “reinforce discriminatory attitudes,” or “favour one group over another (for example, campus-based over distance learning students)” (p. 35). Data-supported wellbeing amplifies both the potential benefits and the risks. Different student experiences and expectations may increase the risk of accidentally discriminatory data; inappropriate assessments or offers of support can cause serious harm (UHI, 2017). The GDPR provides a tool to explore and resolve these high-stakes balances: the Data Protection Impact Assessment (DPIA; Art. 35 and Rec. 90–94).

3.6.1. Data Protection Impact Assessments

DPIAs provide a framework for discussing the benefits and risks of proposed processing. These must be assessed from the perspective of the individuals whose data are processed, not that of the institution, and must consider risks to all rights and freedoms — including non-discrimination, free speech, and free assembly — not just privacy and data protection (Rec. 75). This matches Slater’s requirements to Enable Positive Interventions and Minimize Adverse Impacts.

The DPIA process should involve consultation with affected individuals and groups, discussing what the benefits and risks might be, how the former might be maximized and the latter mitigated, leading to a decision on whether the processing is justified (Art. 35(9)). This process, and the publication and maintenance of the resulting DPIA document, can build trust in the institution’s wellbeing processes, and that its actions are necessary and proportionate (see 3.1 above). For processing involving health data, a DPIA is likely to be required by law (Information Commissioner’s Office, 2019a). If any high risks remain unmitigated but the institution still wishes to proceed, it must seek approval from its national Data Protection Authority, no matter how great the benefits (Art. 36).

Many general processes and templates exist for conducting DPIAs (e.g., Information Commissioner’s Office, 2019a; CNIL, n.d.; UCISA, n.d.). Given the clear relevance to data-supported wellbeing activities, we developed a specific DPIA template as part of our toolkit (Cormack & Reeve, 2020, Annex A). This suggests likely risks and mitigations, and provides a guide to identifying appropriate lawful bases for different processes.

3.7. Stewardship

Sclater's stewardship section "remind[s] institutions of their responsibilities to look after student data carefully" (2016, p. 36). This remains true for wellbeing — see, in particular the GDPR's Accountability Principle (Art. 5(2)) — but wider impacts, including on staff and society, also need to be considered.

Whereas learning analytics is clearly the role of teaching institutions, wellbeing support may cross the remits of several public and private sector bodies. When working in areas of "substantial public interest," GDPR Article 9(2)(g) requires each body to stay within its remit as defined by "Union or Member State law" including, according to the UK Information Commissioner's Office (2019c), "clear common law tasks, functions or powers as well as those set out in the statute or statutory guidance." Papers from AMOSSHE (2015), Universities UK (2015), and Pinsent Masons (Sladdin, 2018) discuss the extent of universities' duty of care to students and the limits beyond which that duty and this GDPR Article indicate that they should call on others. Where remits may overlap, institutions should work with relevant bodies to ensure this produces synergy rather than duplication. Calling on external services may help avoid conflicts of duty, privacy, or trust for individuals within the institution.

DPA18 Section 19(2) warns against processing "likely to cause substantial damage or substantial distress," so institutions should consider wider impacts — on staff, groups, and society — not just on individual students. Since the aim is to improve wellbeing, we should be particularly alert to less concrete forms of harm, such as feelings of "surveillance" or "creepiness" (Jutting, 2016). GDPR Article 22(4) is particularly concerned about automated decisions using special category data. In wellbeing, data and algorithms should support, not replace, human decision makers.

4. Conclusion

Our approach received a considerable boost in July 2019 when the project was accepted into the UK Data Protection Regulator's first Regulatory Sandbox. Not only did this show that the Regulator was willing — despite the high-risk nature of the processing involved — to contemplate that data-supported wellbeing might be lawful, it also provided an opportunity for extensive in-depth discussion with their expert staff. This helped us to understand the more opaque GDPR and DPA18 provisions, and especially to develop wellbeing-focused tools for Purpose Compatibility and Data Protection Impact Assessments (Information Commissioner's Office, 2020).

These tools, described in 3.2.1 and 3.6.1 above, were published as part of Jisc's Wellbeing Analytics Code of Practice in July 2020 (Cormack & Reeve, 2020). This retains the structure of Sclater and Bailey's Learning Analytics Code (2018) and invokes its content, since our review of the GDPR confirmed that structure, and nearly all the content, is a sound basis for investigating and implementing data-supported wellbeing. The only significant change is discussed in Access (3.5 above). However, as we had hoped, our review also identified many additional safeguards, conditions, and tools to apply when using data for wellbeing, rather than tutorial, purposes. These became the main text of our Wellbeing Code. Since our Code builds on the text, governance, and practical arrangements of its predecessor, institutions that have already implemented the Learning Analytics Code have a good basis for exploring whether and how data can support wellbeing. The linked Codes should also help bridge the gap identified by Ahern (2020) between institutional Learning Analytics and Wellbeing policies.

Following publication, we presented the Code at meetings of Data Protection Officers, Learning Analytics practitioners, and student services. It was shared with Universities UK, the National Union of Students, and a civil society organization working in the area. All felt it should significantly help institutions seeking a framework for safe exploration of data-supported wellbeing. Building on the Learning Analytics Code — which has already been widely adopted (e.g., University of Leeds, 2019) — gave confidence that the Wellbeing Code would be practical for universities and colleges to use. One multi-partner project immediately used the DPIA framework to assess their own plans.

The outbreak of COVID-19 led most institutions to focus on reducing known systemic causes of wellbeing issues — for example, using the "whole university" approach (Universities UK, 2020) — rather than exploring new ways to identify and support individual problems. However, we expect our Code — like the Learning Analytics Code on which it is based — to provide a trustworthy framework for such developments, as and when they become possible.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article

Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

Acknowledgments

We are grateful to the many wellbeing, privacy, and data protection experts who gave feedback on the ideas, approach and results of our work. Key documents from this study are available in the following internet archives:

- Cormack, A. N., & Reeve, D. (2020). Code of practice for wellbeing and mental health analytics. <https://web.archive.org/web/20211028163748/https://www.jisc.ac.uk/guides/code-of-practice-for-wellbeing-and-mental-health-analytics>
- Information Commissioner's Office. (2020, June). Regulatory sandbox final report: Jisc. https://web.archive.nationalarchives.gov.uk/ukgwa/*/https://ico.org.uk/media/for-organisations/documents/2618023/jisc-regulatory-sandbox-final-report.pdf
- Sclater, N., & Bailey, P. (2018). Code of practice for learning analytics. <https://web.archive.org/web/20211125143752/https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics>.

References

- Abdi, S., Khosravi, H., Sadiq, S., & Gašević, D. (2020). Complementing educational recommender systems with open learner models. *Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK '20)*, 23–27 March 2020, Frankfurt, Germany (pp. 360–365). ACM. <https://doi.org/10.1145/3375462.3375520>
- Ahern, S. J. (2017, November 17). Learning analytics as a tool for supporting student wellbeing: Identifying student mental ill-health. University College of London. <https://blogs.ucl.ac.uk/digital-education/2017/11/20/learning-analytics-as-a-tool-for-supporting-student-wellbeing-identifying-student-mental-ill-health/>
- Ahern, S. J. (2018). The potential and pitfalls of learning analytics as a tool for supporting student wellbeing. *Journal of Learning and Teaching in Higher Education*, 1(2), 165–172. <https://journals.le.ac.uk/ojs1/index.php/jlthe/article/view/2812>
- Ahern, S. J. (2020). Making a #Stepchange? Investigating the alignment of learning analytics and student wellbeing in United Kingdom higher education institutions. *Frontiers in Education* 5:531424. <https://doi.org/10.3389/educ.2020.531424>
- AMOSSHE. (2015, May 29). Where's the line? How far should universities go in providing duty of care for their students? <https://www.amosshe.org.uk/futures-duty-of-care-2015>
- Article 29 Working Party. (2007). Working document on the processing of personal data relating to health in electronic health records (EHR) (Report no. WP131). https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp131_en.pdf
- Article 29 Working Party. (2014). Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC (Report no. WP 217). https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf
- BBC. (2019, October 5). Stressed students “seeking help” amid fears for academic record. <https://www.bbc.co.uk/news/uk-england-bristol-49852548>
- Binns, R. (2019, October 15). Enabling access, erasure, and rectification rights in AI systems. ICO AI blog. <https://ico.org.uk/about-the-ico/news-and-events/ai-blog-enabling-access-erasure-and-rectification-rights-in-ai-systems/>
- CNIL. (n.d.). Privacy impact assessment (PIA). <https://www.cnil.fr/en/privacy-impact-assessment-pia>
- Cormack, A. N. (2016). Downstream consent: A better legal framework for big data. *Journal of Information Rights, Policy and Practice*, 1(1). <http://doi.org/10.21039/irpandp.v1i1.9>
- Cormack, A. N., & Reeve, D. (2020). Code of practice for wellbeing and mental health analytics. Jisc. <https://www.jisc.ac.uk/guides/code-of-practice-for-wellbeing-and-mental-health-analytics>
- Corrin, L., Kennedy, G., French, S., Buckingham Shum, S., Kitto, K., Pardo, A., West, D., Mirriahi, N., & Colvin, C. (2019). The ethics of learning analytics in Australian higher education: A discussion paper. University of Melbourne. <https://melbourne-cshe.unimelb.edu.au/research/research-projects/edutech/the-ethical-use-of-learning-analytics>
- D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. MIT Press.
- Education Policy Institute. (2018, September 10). Prevalence of mental health issues within the student aged population. <https://epi.org.uk/publications-and-research/prevalence-of-mental-health-issues-within-the-student-aged-population/>
- Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a “right to an explanation” is probably not the remedy you are looking for. *Duke Law & Technology Review*, 16, 18–84. <https://dltr.law.duke.edu/2017/12/04/slave-to-the-algorithm-why-a-right-to-an-explanation-is-probably-not-the-remedy-you-are-looking-for/>
- Ferguson, R. (2019). Ethical challenges for learning analytics. *Journal of Learning Analytics*, 6(3), 25–30. <https://doi.org/10.18608/jla.2019.63.5>

- Ferguson, R., Clow, D., Griffiths, D., & Brasher, A. (2019). Moving forward with learning analytics: Expert views. *Journal of Learning Analytics*, 6(3), 43–59. <https://doi.org/10.18608/jla.2019.63.8>
- Gil, N. (2015, December 14). Majority of students experience mental health issues, says NUS survey. *The Guardian*. <https://www.theguardian.com/education/2015/dec/14/majority-of-students-experience-mental-health-issues-says-nus-survey>
- Griffiths, D., Drachsler, H., Kickmeier-Rust, M., Steiner, C., Hoel, T., & Greller, W. (2016). Is privacy a show-stopper for learning analytics? A review of current issues and solutions. *Learning Analytics Review*, 6. <http://www.laceproject.eu/learning-analytics-review/is-privacy-a-show-stopper/>
- Hall, M. (2019). Student wellbeing and mental health: The opportunities in learning analytics. Jisc. <http://repository.jisc.ac.uk/6916/1/student-wellbeing-and-mental-health-the-opportunities-in-learning-analytics.pdf>
- Hughes, G., & Spanner, L. (2019). The university mental health charter. StudentMinds. https://www.studentminds.org.uk/uploads/3/7/8/4/3784584/191208_umhc_artwork.pdf
- Information Commissioner's Office. (n.d.). How do we document our processing activities? Guide to the GDPR. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/documentation/how-do-we-document-our-processing-activities/>
- Information Commissioner's Office. (2019a, May 22). Data protection impact assessments. Guide to the General Data Protection Regulation (GDPR). <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/>
- Information Commissioner's Office. (2019b, May 22). How do we apply legitimate interests in practice? Guide to the General Data Protection Regulation (GDPR). <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/>
- Information Commissioner's Office. (2019c, May 22). Public task. Guide to the Data Protection Regulation. <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/>
- Information Commissioner's Office. (2019d, May 22). Special category data. Guide to the General Data Protection Regulation (GDPR). <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/>
- Information Commissioner's Office. (2019e, May 22). What are the substantial public interest conditions? Guide to the General Data Protection Regulation (GDPR). <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/special-category-data/what-are-the-substantial-public-interest-conditions/>
- Information Commissioner's Office. (2020, June). Regulatory sandbox final report: Jisc. <https://ico.org.uk/media/for-organisations/documents/2618023/jisc-regulatory-sandbox-final-report.pdf>
- Inkster, B., & Digital Mental Health Data Insights Group. (2021). Early warning signs of a mental health tsunami: A coordinated response to gather initial data insights from multiple digital services providers. *Frontiers in Digital Health* (10 February 2021). <https://doi.org/10.3389/fdgth.2020.578902>
- Jisc. (n.d.). Who we are and what we do. <https://www.jisc.ac.uk/about/who-we-are-and-what-we-do>
- Jisc. (2019). Horizons report on emerging technologies and education. <https://www.jisc.ac.uk/reports/horizons-report-emerging-technologies-and-the-mental-health-challenge>
- Jones, L. (2019). Using LA to inform student wellbeing and have better conversations with students. Paper presentation to the Jisc LA Cymru User Group, 3 April 2019.
- Jutting, C. (2016, August 3). Universities are tracking their students. Is it clever or creepy? *The Guardian*. <https://www.theguardian.com/higher-education-network/2016/aug/03/learning-analytics-universities-data-track-students>
- Kaliisa, R., Kluge, A., & Mørch, A. (2020). Combining checkpoint and process learning analytics to support learning design decisions in blended learning environments. *Journal of Learning Analytics*, 7(3), 33–47. <https://doi.org/10.18608/jla.2020.73.4>
- Kitchener, S. (2016, January 12). OccupEye sensors: A sinister exercise in Big Brother-style management or a 21st-century way to monitor workers' needs? *Independent*. <https://www.independent.co.uk/news/media/occupeye-sensors-sinister-exercise-big-brother-style-management-or-21st-century-way-monitor-workers-needs-a6808281.html>
- Lang, C., Woo, C., & Sinclair, J. (2020). Quantifying data sensitivity: Precise demonstration of care when building predictive models. *Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK '20)*, 23–27 March 2020, Frankfurt, Germany (pp. 655–664). ACM. <https://doi.org/10.1145/3375462.3375506>

- Mack, D., DaSilva, A., Rogers, C., Hedlund, E., Murphy, E., Vojdanovski, V., Plomp, J., Wang, W., Nepal, S., Holtzheimer, P., Wagner, D., Jacobson, N., Meyer, M., Campbell, A., & Huckins, J. (2021). Mental health and behavior of college students during the COVID-19 pandemic: Longitudinal mobile smartphone and ecological momentary assessment study. *Journal of Medical Internet Research*, 23(6). <https://doi.org/10.2196/28892>
- McKie, A. (2019, April 14). Could student mental health apps be doing more harm than good? *Times Higher Education*. <https://www.timeshighereducation.com/news/could-student-mental-health-apps-be-doing-more-harm-good>
- Mulenda, H. (2020, September 4). Harnessing the power of technology can help vulnerable students. *Times Higher Education*. <https://www.timeshighereducation.com/opinion/harnessing-power-technology-can-help-vulnerable-students>
- Neves, J., & Hillman, N. (2017). *Student academic experience survey*. Higher Education Authority/Higher Education Policy Institute. <https://www.hepi.ac.uk/wp-content/uploads/2017/06/2017-Student-Academic-Experience-Survey-Final-Report.pdf>
- National Union of Students. (2015, August). Learning analytics: A guide for students' unions. <https://www.nusconnect.org.uk/resources/learning-analytics-a-guide-for-students-unions>
- Office for National Statistics. (2021, March). Coronavirus and higher education students: England, 19 February to 1 March 2021. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandhighereducationstudents/19februaryto1march2021>
- Perkins, D., Gray, C., Ritsos, P., & Kuncheva, L. (2020). Jisc/Bangor University learning analytics project summary & case study. Jisc. <http://research.academicanalytical.com/jspui/handle/1471/31>
- Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency and learning analytics: An exploration. *Journal of Learning Analytics*, 3(1), 159–182. <https://doi.org/10.18608/jla.2016.31.10>
- Sclater, N. (2016). Developing a code of practice for learning analytics. *Journal of Learning Analytics*, 3(1), 16–42. <https://doi.org/10.18608/jla.2016.31.3>
- Sclater, N., & Bailey, P. (2018). Code of practice for learning analytics. Jisc. <https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics>
- Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics*, 6(3), 11–19. <https://doi.org/10.18608/jla.2019.63.3>
- Sladdin, J. (2018, Oct 17). Duty to care for student mental health has legal implications for universities. Out-Law. <https://www.pinsentmasons.com/out-law/analysis/duty-to-care-student-mental-health-legal-implications-universities>
- Slade, S., & Prinsloo, P. (2014). Student perspectives on the use of their data: Between intrusion, surveillance and care. *Challenges for Research into Open & Distance Learning: Doing Things Better – Doing Better Things*, 291–300. <https://www.eden-online.org/proc-2485/index.php/PROC/article/view/1326>
- Slade S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. *Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK '20)*, 23–27 March 2020, Frankfurt, Germany (pp. 235–244). ACM. <https://doi.org/10.1145/3303772.3303796>
- Thorley, C. (2017). *Not by degrees: Improving student mental health in the UK's universities*. IPPR. <https://www.ippr.org/research/publications/not-by-degrees>
- UCISA. (n.d.). *Privacy impact assessment toolkit*. <https://www.ucisa.ac.uk/PIAToolkit>
- UHI. (2017). *Mental health conditions toolkit*. <https://staffresources.uhi.ac.uk/mhc/>
- Universities UK. (2015). *Student mental wellbeing in higher education: Good practice guide*. <https://www.m25lib.ac.uk/wp-content/uploads/2021/02/student-mental-wellbeing-in-he.pdf>
- Universities UK. (2020). Stepchange: Mentally healthy universities. <https://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2020/uuk-stepchange-mhu.pdf>
- Universities UK. (2021). Minding our future: Starting a conversation about the support of student mental health. <https://www.universitiesuk.ac.uk/what-we-do/policy-and-research/publications/minding-our-future-starting-conversation>
- University of Leeds. (2019, May 24). Code of practice on learning analytics. https://forstaff.leeds.ac.uk/info/30402/university_news/1882/code_of_practice_on_learning_analytics
- University of South Wales. (n.d.). Progression advice team. <https://progression.southwales.ac.uk/>
- Verbert, K., Ochoa, X., De Croon, R., Dourado, R., & De Laet, T. (2020). Learning analytics dashboards: The past, the present and the future. *Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK '20)*, 23–27 March 2020, Frankfurt, Germany (pp. 35–40). ACM. <https://doi.org/10.1145/3375462.3375504>
- Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, 2019(2), 494–620. <https://doi.org/10.7916/cblr.v2019i2.3424>

Woodward, M. (2021). 16 countries with GDPR-like privacy laws. <https://securityscorecard.com/blog/countries-with-gdpr-like-data-privacy-laws>