



HAL
open science

A Multi-level Analysis of Mistrust/Trust Formation in Algorithmic Grading

Stephen Jackson, Niki Panteli

► **To cite this version:**

Stephen Jackson, Niki Panteli. A Multi-level Analysis of Mistrust/Trust Formation in Algorithmic Grading. 20th Conference on e-Business, e-Services and e-Society (I3E), Sep 2021, Galway, Ireland. pp.737-743, 10.1007/978-3-030-85447-8_61 . hal-03648163

HAL Id: hal-03648163

<https://inria.hal.science/hal-03648163v1>

Submitted on 21 Apr 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

A Multi-Level Analysis of Mistrust/Trust Formation in Algorithmic Grading

Stephen Jackson¹ and Niki Panteli²

¹ Ontario Tech University, Oshawa, ON L1G 0C5, Canada
stephen.jackson@uoit.ca

² Royal Holloway, Egham, Surrey, TW20 0EX, United Kingdom & Norwegian University of Science and Technology, Trondheim, Norway
niki.panteli@rhul.ac.uk

Abstract

While the concept of trust continues to grow in importance among information systems (IS) researchers and practitioners, an investigation of mistrust/trust formation in algorithmic grading across multiple levels of analysis has so far been under researched. This paper proposes a multi-level model for analyzing the formation of mistrust/trust in algorithmic grading. More specifically, the model examines multiple levels at play by considering how top-down forces may stimulate mistrust/trust at lower levels, but also how lower-level activity can influence mistrust/trust formation at higher levels. We briefly illustrate how the model can be applied by drawing on the case of the Advanced Level student fiasco in the United Kingdom (UK) that came to head during August 2020, whereby an algorithm was used to determine student grades. Although the paper positions trust as a multifaceted concept, it also acknowledges the importance of researchers to be mindful of issues pertaining to emergence, duality, context, and time.

Keywords: Trust, Mistrust, Algorithmic Grading, Multi-Level Analysis, Multi-Level Theory.

1 Introduction

There has been growing interest in the development and use of computerized algorithmic grading within the educational sector (Kapudi et al., 2021). Put simply, algorithmic grading can be defined as a set of rules or procedures undertaken by computerized systems that can be used to perform calculations or problem solving to assist with student grade prediction. Proponents have acknowledged the potential benefits of artificial intelligence (AI) and algorithms in education for speeding up marking, standardizing the grading process, driving through improvements in efficiencies and forecasting, as well as freeing up instructors marking time to concentrate on other value-added activities (Goel & Joyner, 2017; Khare et al., 2018).

On the other hand, concerns have also been raised. By placing increased control in a set of computerized rules and norms may lead to tutors doubting their own professional expertise, judgments and opinions, as well as feeling somehow negatively constrained, if not alienated (Guskey & Jung, 2016; Kolchenko, 2018). As algorithmic grading develops through time, particularly with advances in AI, this may shape and, in cases, raise the expectations of tutors into believing that these new breeds of applications can provide a more reliable, accurate and fairer tool for marking over traditional methods. Additionally, the increased reliance on computerized grading

tools may ignore or downplay other important contextual factors that can influence student performance at particular points in time. Concerns have also been raised in relation to the overall effectiveness of AI-based grading algorithms in the electronic age. Parsons (2020), for instance, reported how US students were able to outwit the AI algorithm used to assign assessment grades. Realizing that the algorithm was merely scanning a series of keywords based on the topic of the assessment, students were able to “game the system” by inserting keywords, rather than provide a detailed and comprehensive answer.

While the use of algorithms has the potential to bring educational benefits, the extent to which the educational sector can truly harness AI and algorithms for marking and other educational purposes is still not clear and evolving (Popenici & Kerr, 2017; Schiff, 2021). One important issue that lies at the very heart of computerized applications, affecting how they are successfully used and deployed, is trust. Li et al. (2008, p.41), based on the work of Mayer et al.’s (1995), define trust as the “willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party”. Siau and Wang (2018) acknowledged that trust is an essential element in building relations and for nurturing user acceptance. The authors stress that in the case of AI “both initial trust formation and continuous trust development deserve special attention” (Siau & Wang, 2018, p.47). In a similar vein, Gille et al. (2020p.1) note “trust is a critical determinant of the successful adoption of AI”.

Although the physical aspects of AI-based technologies, including how it performs and functions, can affect trust formation (Sethumadhavan, 2019), what must also be considered is how those responsible for developing and deploying AI can influence the trust creation process (Siau & Wang, 2018). A high degree of trust, as a confident and optimistic anticipation in the actions and behavior of another stakeholder, can facilitate integrity and transparency. In conditions of vagueness or when operating under a situation of incomplete information and knowledge, the establishment of trust between stakeholders becomes an ever-important element for reducing complexity and nurturing openness (Adler, 2001; Currall & Judge, 1995; Fukuyama, 1996; Inkpen, 1996). The importance of trust is further accentuated by its absence and its replacement with mistrust which weakens and jeopardizes relationships (Luhmann, 1979). Mistrust can have an unconstructive impact on building a culture conducive to AI use, promoting values of suspicion and protectiveness (Jackson & Panteli, 2020). Furthermore, it has been acknowledged that if trust is not evident, this can make the use and adoption of technology more difficult to attain (Jackson, 2011; Mohr & Walter, 2019; Sharma & Sharma, 2019; Wagner & Newell, 2006).

Regardless of the importance of trust in the adoption and use of AI and technology in general, our understanding of mistrust/trust in relation to the study of algorithmic grading from a multi-level analysis has been limited. It is our position that with the increasing popularity of AI for educational purposes, this issue requires further attention. This paper acknowledges that in order for researchers to understand how mistrust/trust influences algorithmic grading uptake and use, among other forms of AI, the focus of attention needs to shift from solely concentrating on just one level to also understanding the multi-faceted and dynamic nature of trust across multiple levels.

In what follows, we present the case of the Advanced (A) Level AI application to determine student grades in August 2020 in the United Kingdom. Drawing on this case, we introduce a model based on multi-level theory.

2 The Case of the A-Level Exam Algorithm

In the year 2020, the COVID-19 pandemic caused an unprecedented disruption to the educational sector worldwide, contributing to the redesign of learning and assessment systems (Pappas & Giannakos, 2021). Faced with a tough decision, The Office of Qualifications and Examinations Regulation (Ofqual) in the UK had to decide how best to predict student A-Level grades due to the problems caused by the COVID-19 pandemic that prevented students from physically sitting their exams (Katwala, 2020).

A-Level grades play a major part in determining university admissions, as well as direct routes into employment and apprenticeships. As the reliance on student grade predictions by teachers may have run the risk of grade inflation, it was decided to use an algorithm to calculate grades. According to the BBC (2020) “Teachers were asked to supply for each pupil for every subject: an estimated grade and a ranking compared with every other pupil at the school within that same estimated grade. These were put through an algorithm, which also factored in the school's performances in each subject over the previous three years. The idea was that the grades this year - even without exams - would be consistent with how schools had done in the past”.

However, the algorithm wreaked havoc in that it downgraded tens of thousands of students. When the results were announced in August 2020, almost 40% of grades in Wales, Northern Ireland and England were lower than the predictions provided by teachers. What became apparent was that the algorithm affected students from state schools more than those from private schools. In other words, students from poorer backgrounds were more likely to receive lower grades than those from areas that are more affluent. A major reason for this was that the algorithm took into consideration the historical performance of the school. However, as put by Elbanna and Engesmo (2020) “...data about the past doesn't necessarily help you make adequate decisions about the present or the future. It blocks any chances of change and development – like when a school improves its teaching or one year group of students performs better than their peers in previous years”. Therefore, a good student who was from a school that underachieved may likely have his/her grades downgraded. Even though the UK government initially presented the AI system as “robust”, within a few days and following public uproar, it made a u-turn and abolished the system reversing to teacher assessment instead (BBC, 2020).

3 Conceptual Multi-Level Model and Application

The proposed model is based on multi-level theory (e.g., Erez & Gati, 2004; Klein & Kozlowski, 2000; Lumineau & Schilke, 2018). The model was formulated out of the concern that researchers tend to study one level, rather than accounting for multiple levels. A key assumption is that considering the phenomena of interest from just one level, can potentially, although not always, lead to a restrictive view. As trust is fundamentally about interrelated social relations, it is difficult not to consider trust as consisting of social relations influenced by other social relations (levels).

For this reason, a multi-level model which considers government, organizational, group and individual factors across different stages of AI development and use, is useful as it can be applied to understand the multi-faceted nature of mistrust/trust, particularly the dynamic interplay between macro and micro levels. The government level of analysis considers the structure and nature of the political/government system and how this can influence the AI trust enactment process. At the organizational level, the focus is on the way company size, culture, values and practices foster trust or mistrust in AI. At the group level, individuals may group together to share common values, beliefs and assumptions regarding trust and AI. Trust at the individual layer reflects the values of oneself. This may include the personal values, beliefs, attitudes and prejudices which users have towards AI. An overview of the proposed model is shown in figure 1.

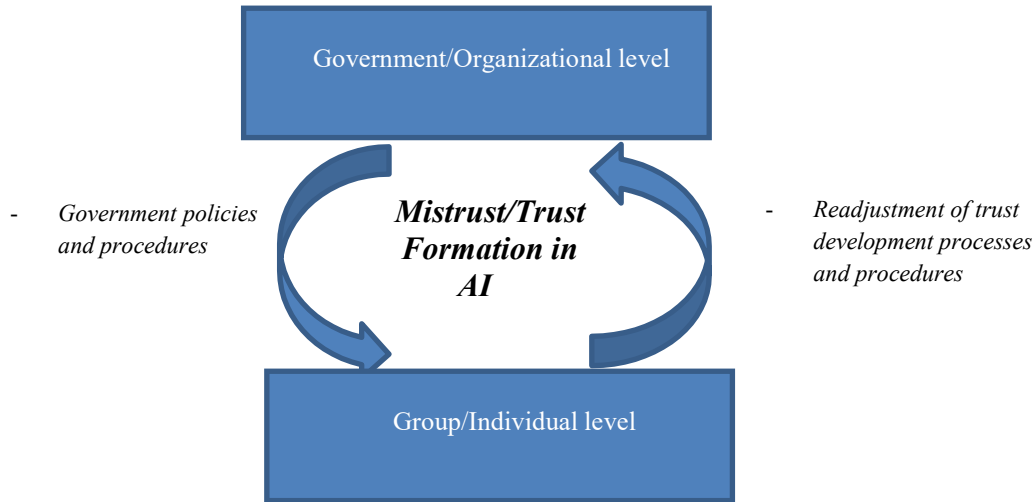


Fig. 1. Conceptual model

Higher-level activity can shape and influence trust formation at lower levels, but lower levels may in turn influence trust development at higher levels. The influence at various levels is best considered as dynamic and changing. Not only does the proposed model bridge macro and micro levels, but also the situational context that gives rise to trust/mistrust should not be neglected. Our position is that these influences should take place as part of a co-creative effort rather than at separate stages as it was the case with the A-level AI example.

Drawing on figure 1, top-down forces at play e.g., government policies, procedures and practices can influence the formation of mistrust/trust among lower-level activity, enabling or constraining the actions of groups and individuals. In the case of the A-Level grading controversy, the use of the Ofqual algorithm to predict grades, the inclusion of the schools past performance over 3 years, as well as not basing grades solely on teacher predicted grades, reinforced a constraining environment at lower levels (Coughlan, 2020; Kolkman, 2020). Many students, because of their grades being downgraded, lost their university place and blamed the government for playing with their future. Students, teachers, parents and politicians united in opposition, as exemplified through protests and gatherings, raising fundamental concerns about the fairness, bias and accuracy surrounding the algorithm. The situation that emerged within a very short period of time, at both the group and individual levels, was one of uncertainty and mistrust (Coughlan, 2020).

The model also captures how emergent forms of action and behavior from lower levels can shape and influence changes at the higher level, including trust development processes. While the government initially defended its use of an algorithm to determine grades as being dependable and accurate, in the face of increased scrutiny, it was announced that the algorithm had shortcomings and was not as reliable as first envisioned. Indeed, the Prime Minister Boris Johnston later referred to it as a “mutant algorithm” (Poole, 2020). As opposition to the flawed algorithm grew, the government decided to make a u-turn and it was revealed that grades would be awarded based on teachers’ recommendations (Reynolds, 2020). Although awarding student grades based solely on teacher grade predictions was initially disregarded by Ofqual on the basis of being unfair, lower-level activity reinforced a reversal of norms and policies, as well as readjustment of trust development processes and procedures at the upper-level. More explicitly, rather than

placing unbridled trust in the use of algorithms for grade predictions, trust was now placed in the hands of humans (teachers) – a practice which the government would continue to enact the following year. As Education Secretary Gavin Williamson noted in January 2021 “this year we’re going to put our trust in teachers rather than algorithms” (Sleigh, 2021).

4 Concluding Thoughts

This short paper set out to investigate mistrust/trust formation in the context of algorithmic grading by drawing briefly on the UK A-Level grades debacle in summer 2020. While the study of trust and computerized applications represents a growing and important area of inquiry, research on algorithmic grading and issues pertaining to mistrust/trust are still at a nascent stage and much remains to explore these issues in greater depth. Comprehending how these two areas fuse together is an important endeavor and may reveal unique insights not previously known or explored. As a way of advancing our understanding of trust in relation to the adoption of algorithms, as well as other forms of AI, Robotics and Machine Learning, there is increased need for studies to investigate further its dynamic and multi-faceted nature, how trust emerges across multiple levels, as well as remaining sensitive to emergence, context, dynamism, and change.

References

- Adler, P. (2001). Market, hierarchy and trust: The knowledge economy and the future of capitalism. *Organization Science*, *12*(2), 215-234.
- BBC (2020, August 20). A-levels and GCSEs: How did the exam algorithm work? *BBC News*. Retrieved from <https://www.bbc.co.uk/news/explainers-53807730>.
- Coughlan, S. (2020, August 14). Why did the A-level algorithm say no? *BBC News*. Retrieved from <https://www.bbc.com/news/education-53787203>.
- Currall, S., & Judge, T. (1995). Measuring trust between organization boundary role persons. *Organization Behaviour and Human Decision Processes*, *64*(2), 151-170.
- Elbanna, A., & Engesmo, J. (2020, August 19). A-level results: why algorithms get things so wrong – and what we can do to fix them. *The Conversation*. Retrieved from <https://theconversation.com/a-level-results-why-algorithms-get-things-so-wrong-and-what-we-can-do-to-fix-them-142879>
- Erez, M., & Gati, E. (2004). A dynamic, multi-level model of culture: from the micro level of the individual to the macro level of a global culture. *Applied Psychology: An International Review*, *53*(4), 583-598.
- Fukuyama, F. (1996). *Trust: the social virtues and the creation of prosperity*. London, Penguin.
- Gille, F., Jobin, A., & Ienca, M. (2020). What we talk about when we talk about trust: theory of trust for AI in healthcare. *Intelligence-Based Medicine*, *1*, 100001.
- Goel, A. K., & Joyner, D. A. (2017). Using AI to teach AI: lessons from an online AI class. *AI Magazine*, *38*(2), 48-59.
- Guskey, T. R., & Jung, L. A. (2016). Grading: Why you should trust your judgment. *Educational Leadership*, *73*(7), 50.
- Inkpen, A. (1996). Creating knowledge through collaboration. *California Management Review*, *39*(1), 123-140.
- Jackson, S. (2011). Organizational culture and information systems adoption: a three-perspective approach. *Information and Organization*, *21*(2), 57-83.
- Jackson, S. & Panteli, N. (2020, September 20). Trust in the era of artificial intelligence: a multi-layer analysis. *34th Annual British Academy of Management Conference*. Manchester, England.

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: a systematic mapping of the literature. *Computers and Education: Artificial Intelligence*, 100017.

Katwala, M. (2020, August 15). An algorithm determined UK students' grades. Chaos ensued. *Wired*. Retrieved from <https://www.wired.com/story/an-algorithm-determined-uk-students-grades-chaos-ensued/>

Khare, K., Stewart, B., & Khare, A. (2018). Artificial intelligence and the student experience: an institutional perspective. *IAFOR Journal of Education*, 6(3), 63-78.

Klein, K., & Kozlowski, S. (2000). *Multilevel theory, research, and methods in organizations: Foundations, Extensions and New Directions*. San Francisco, CA Jossey-Bass.

Kolchenko, V. (2018). Can modern AI replace teachers? Not so fast! Artificial intelligence and adaptive learning: personalized education in the AI age. *HAPS Educator*, 22(3), 249-252.

Kolkman, D. (2020, August 26). "F**k the algorithm"? What the world can learn from the UK's A-level grading fiasco. *The London School of Economics Blog*. Retrieved from <https://blogs.lse.ac.uk/impactofsocialsciences/2020/08/26/fk-the-algorithm-what-the-world-can-learn-from-the-uks-a-level-grading-fiasco/#comments>.

Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems*, 17(1), 39-71.

Luhmann, N. (1979). *Trust and Power*. London: John Wiley and Sons.

Lumineau, F., & Schilke, O. (2018). Trust development across levels of analysis: an embedded-agency perspective. *Journal of Trust Research*, 8(2), 238-248.

Mayer, R. Davis, J. & Schorman, F. (1995). An integrative model of organizational trust. *Academy of Management Journal*, 20(3), 709-734.

Mohr, H. & Walter, Z. (2019). Formation of consumers' perceived information security: examining the transfer of trust in online retailers. *Information Systems Frontiers*, 21, 1231-1250.

Pappas, I. O., & Giannakos, M. N. (2021). Rethinking learning design in IT education during a pandemic. *Frontiers in Education*, 6, 103.

Parsons, J. (2020, September 4). Students figure out how to cheat AI grading algorithm. *Metro*. Retrieved from <https://metro.co.uk/2020/09/04/students-figure-out-how-to-cheat-ai-grading-algorithm-13222401/>.

Poole, S. (2020, September 3). 'Mutant algorithm': boring B-movie or another excuse from Boris Johnson? *The Guardian*. Retrieved from <https://www.theguardian.com/books/2020/sep/03/mutant-algorithm-boring-b-movie-or-another-excuse-from-boris-johnson>.

Popenici, S. A., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12(1), 1-13.

Reynolds, E. (2020, August 17). UK government makes dramatic exam results U-turn after national outcry. *CNN News*. Retrieved from <https://www.cnn.com/2020/08/17/europe/uk-school-exams-coronavirus-gbr-intl/index.html>.

Schiff, D. (2021). Out of the laboratory and into the classroom: the future of artificial intelligence in education. *AI & Society*, 36(1), 331-348.

Sethumadhavan, A. (2019). Trust in artificial intelligence. *Ergonomics in Design*, 27(2), 34-34.

Sharma, S. & Sharma, M. (2019). Examining the role of trust and quality dimensions in the actual usage of mobile banking services: an empirical investigation. *International Journal of Information Management*, 44, 65-75.

Siau, K. & Wang, W. (2018). Building trust in artificial intelligence, machine learning and robotics. *Cutter Business Technology Journal*, 31(2), 47-53.

Sleigh, S. (2021, January 6). Exams in England will be replaced by school-based assessments, says Gavin Williamson. *Evening Standard*. Retrieved from <https://www.standard.co.uk/news/education/gavin-williamson-a-levels-gsces-remote-learning-b719481.html>.

Wagner, E. & Newell, S. (2006). Repairing ERP: Producing social order to create a working information system. *Journal of Applied Behavioural Research*, 42(1), 40-57.