Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Repurposing Existing Data Towards Institutional Learning Analytics: A Review of Outcome-mapping Data of HEIs in India

Debarun SARKAR[a*] & Anitha KURUP[b]
[a]*Vidyashilp Research Centre, Vidyashilp University, Bengaluru, India*
[b]*Education Program, National Institute of Advanced Studies, Bengaluru, India; Research and Innovation Council, Vidyashilp University, Bengaluru, India*
*debarun.sarkar@vidyashilp.edu.in

Abstract: The paper addresses the possibility of repurposing existing data structures in Indian higher education institutes (HEIs) for the deployment of institutional learning analytics (LA). To that end, the paper critically reviews the kind of data that HEIs in India already generate for quality control and accreditation by the National Assessment and Accreditation Council. It argues that, albeit with caveats, existing data structures of programme and course outcomes maintained by HEIs can be repurposed toward institutional LA deployment. A significant amount of work must be done to flesh out the learning theories that will inform the data structures. Centring learning outcomes for LA deployment can help address the various existing critiques of LA. At the same time, outcome-based approaches risk expunging unexpected, abstract and social learnings and risk reproducing managerialist approaches to education.

Keywords: creative data sourcing, institutional learning analytics, quality assurance in higher education, India

1. Introduction

Institutional deployment of learning analytics is at a nascent and embryonic stage in India. This trend reflects the broader lack of LA research and practice in the global south. Various possible reasons for the lack of LA research in the global south have been noted including lack of infrastructure, possible under-reporting of experiments, initiatives and publication of research in non-indexed avenues (Guzmán-Valenzuela, Gómez-González, Rojas-Murphy Tagle, & Lorca-Vyhmeister, 2021).

Guzmán-Valenzuela et al. (2021) limit the infrastructural lack to technologies that facilitate “the promotion of more virtual and blended learning within…universities which, in turn, generates possibilities for data management systems and, thence research" (Guzmán-Valenzuela et al., 2021, p. 12). This equation of LA with online learning has been noted to be one of the significant deficiencies of the existing approaches of LA, as Eradze, Väljataga, & Laanpere (2014, p. 256) note, “most of the tools for gathering the learning analytics data are directed to the closed LMS systems, while most of the learning happens outside the LMS”.

Thinking about the lack of infrastructural lack in the global south, certain authors have called for creative data sourcing strategies (Gašević, 2018). Echoing the principle of data minimization Prinsloo (2018) warns against the rampant call for indiscriminate data collection. The paper builds on such calls for creative data sourcing strategies in the global south to address the viability of repurposing existing quality control and accreditation data which is currently generated by HEIs in India.

HEIs in India, generate a significant amount of data for quality control purposes for the National Assessment and Accreditation Council (NAAC). Though NAAC accreditation remains a voluntary affair, institutions seek NAAC accreditation for consumer assurance in an increasingly privatized higher education sector. The data generated for NAAC retains the methodological limitations of the NAAC-mandated methodology. The paper argues, albeit with caveats, that an evaluation of the outcome-based mapping of programmes and courses and the pedagogic methods can form the ground for an institutional LA deployment. While noting that, the paper also flags the limitations of an outcome-based approach and the limitations of the current data collection methodologies.

2. Existing Work

The paper follows existing literature concerning LA in the global south which has called for "connectedness within local systems and across levels of a system" (Chen & Fan, 2018, p. 41) and comprehending the nature of data that is already being generated and collected (Prinsloo, 2018). These calls have been made amidst the steady adoption of ICTs and the risk of HEIs adopting "commercial providers and platforms" (Prinsloo, 2018, p. 30).

Gašević's call for "creative data sourcing" (2018, p. 8) is a valid call amidst such a context as Prinsloo notes the solution "may not be to harvest more (or different) data" (Prinsloo, 2018, p. 29). We take Gašević's and Prinsloo's calls, for creative data sourcing and mapping the nature of existing data, seriously to grasp the kind of data that HEIs in India are already generating. for quality control, assurance and accreditation purposes.

The NAAC was set up in 1994. Though accreditation by NAAC is not compulsory and remains a voluntary affair as an autonomous body under the aegis of the University Grants Commission (Stella, 2004, 2015). NAAC's accreditation is a much sought-after metric by a third party in an educational landscape which is populated heavily by private and non-state actors. NAAC's process seeks to enable quality control and provides comparative quality indices at the national level that is used by students and parents to make decisions about enrolment into colleges and universities.

The existing NAAC methodology seeks to provide accreditation to HEIs and mandates the institutes to set up an internal quality assurance cell (IQAC) to proactively engage in quality control measures. This internal quality assurance mechanism is complemented by a peer-review of the institutional reports submitted during the accreditation process whose scores remain valid for five years. The existing NAAC methodology remains skewed towards institutional accreditation over departmental accreditation.

In a recent whitepaper published by NAAC, Patwardhan et al. (2022) argue for a massive overhaul of the existing NAAC methodology away from an input-based system to an output-based system. They argue for a "shift from the current fixed time-point data entry…and peer team visit based summative assessment system…to capture real-time data and continuous assessment of education quality and expected outcomes" (Patwardhan et al., 2022, p. 58). Patwardhan et al. (2022) also call for a refocusing on learning outcomes (LOs) rather than focusing on other proxies, as they note "quality of teaching-learning is currently assessed by proxy parameters like teacher-student ratio, number of PhD holders in the Faculty, number of books in the library, and so on" (Patwardhan et al., 2022, p. 37).

The speed and the specifics of the transition remain uncertain at the moment. Hence, the paper follows Gašević's call for "creative data sourcing" to ascertain if the existing data being generated for the NAAC can be repurposed for the institutional deployment of LA while centring LOs. The paper also builds on existing works which have argued for existing data being generated for the NAAC toward output-based education (Amirtharaj, Chandrasekaran, Thirumoorthy, & Muneeswaran, 2022).

3. Methodology

To gain familiarity with the existing data structures that HEIs generate and manage in India, the standardized data template used by HEIs to submit data to NAAC was referred to. Following this, self-study reports (SSR) submitted by HEIs to the NAAC were referred to critically assess the granularity of data concerning learning processes and outcomes. Additional documents and data submitted with the SSRs were referred to understand precisely the nature of data currently present with HEIs concerning learning outcomes and processes.

4. Repurposing NAAC data for institutional LA deployment

4.1 Nature of Data Generated for Accreditation by the NAAC

HEIs in India generate a significant amount of data for the accreditation process for the NAAC. Institutions which have gone through multiple cycles of accreditation have an IQAC in place which manages the generation of data for the accreditation and quality assurance process. The NAAC publishes a manual (National Assessment and Accreditation Council, 2018) for generating SSRs. The presence of the manual has led to a standardization of reports being generated by HEIs (Stella, 2015). The SSRs include a range of information and data about the HEIs. They often include additional documents which provide further details about the institutions at a granular level.

Various kinds of data and information are present in each of the above criteria of assessment. For heuristic reasons and a lack of space, Table 1 summarises the assessment criteria and their sub-division to provide the reader with a sense of the diverse range of data HEIs in India currently generate. Each of the sub-criteria provides qualitative and quantitative data and information of various kinds. The quality of the information provided often differs depending on the institution.

Table 1. *Some of the criteria of assessment and their subdivisions and their weightage in SSRs relevant for LA deployment.* (National Assessment and Accreditation Council, 2018).

Criteria of assessment	**Sub-criteria**	**Weightage on a scale of 1000 points**
Curricular Aspects	1.1 Curriculum Design and Development	50
	1.2 Academic Flexibility	50
	1.3 Curriculum Enrichment	30
	1.4 Feedback System	20
Teaching-Learning and Evaluation	1.1 Student Enrolment and Profile	10
	2.2 Catering to Student Diversity	20
	2.3 Teaching- Learning Process	20
	2.4 Teacher Profile and Quality	50
	2.5 Evaluation Process and Reforms	40
	2.6 Student Performance and Learning Outcomes	30
	2.7 Student Satisfaction Survey	30
Student Support and Progression	5.1 Student Support	30
	5.2 Student Progression	40
	5.3 Student Participation and Activities	20
	5.4 Alumni Engagement	10

Each of the above sub-divisions provided in Table 2 can form the ground for an institutional LA deployment. One of the persistent critiques of LA has been LA's lack of focus on educational and learning theories (Guzmán-Valenzuela et al., 2021). It is hence, pertinent to begin an institutional framework for LA deployment which starts with a focus on learning processes while at the same time acknowledging the limitations and genealogies of the above data and their rationale.

4.2 An Illustrative Case

In point no. 2.6 (from Table 1), institutions provide detailed information about the curriculum, assessment and learning outcomes methodology of the institution. Institutions are offered an option to share a detailed description of programme outcomes (PO) and course outcomes (CO) and how they are linked to graduate attributes. This detailed mapping of POs and COs then is a task which has already been undertaken by most institutes which have gone through multiple cycles of NAAC accreditation.

Consider a programme of Bachelor of Architecture from an HEI in Bengaluru. The institute provides all of the department's data in the public domain for NAAC accreditation. The Bachelor of Architecture at the HEI under consideration defines the graduate attribute as: "The B Arch Programme intends a deep immersion in an ecosophical perspective of architecture, as part of an inspired understanding of larger discourses: environmental, social, political, artistic and technological".

The data lists the range of programme outcomes divided into three groups, affective, cognitive and psychomotor components. Along with the programme outcomes it maps each course in the programme to the various programme outcomes with the adjacent mastery level mapping. Each course is valued and classified on a three-tier mastery scale: introductory, reinforced and emphasized. For each course outcome and its mapping to programme outcomes the data sets also list out instructional and assessment strategies that are deployed.

Consider for example PO1 of the programme under consideration: "Sensitize students to be socially and environmentally responsible and to work effectively in multi-disciplinary teams within the field of human habitat" as an affective component. In the course Discovering Design from Semester 1, PO1, among others, is classified with mastery level being "introductory". The instructional strategies deployed towards the course outcome are "Studio on Wheels, Lectures, Presentation, Mapping, Film watching, Workshop and Masterclass" while the assessment strategies mobilized include "Pinup reviews, Individual Desk Crits, Group Discussion and Portfolio Submission".

Similarly, the course Specifications, Estimation and Costing of Buildings from Semester 5 aims at a mastery level of 'emphasized' for PO3 which is, "Nurture quality education that enables use and extension of appropriate knowledge for designing built environment" classified as a cognitive component. The instructional strategies include "Lectures, Discussions, Active problem-solving in class" while the assessment strategies include "Assignments, Examinations".

Amirtharaj et al. (2022) provide a view of what the outcome-based approach looks like from the perspective of instructors, teachers, evaluators and administrators in an HEI in India. The COs and POs are quantified for each student at different scales from the individual, across courses to the programme. One particular course, for example, could have more than one CO. As per Amirtharaj et al. (2022) a student is scored not just on the whole course as such but each COs are scored as well. While such granular level mapping of COs and POs is followed, the cycle for assessment of CO and PO attainment is relatively large. They list the frequency of assessment processes as follows:

- "CO–PO mapping—Every year (for each course);
- CO attainment report—Every year (for the three internal tests, assignments and end semester examination);
- PO attainment report—Once for every batch (after course completion);
- PEO attainment report—Once for every batch (after course completion);
- Alumni feedback—After graduation;
- Parents feedback—After graduation of their wards;
- Placement (including campus recruitment) record—Every year; and

- Employer feedback—Occasionally." (Amirtharaj et al., 2022, p. 22)

It is worth noting the fact that such a low frequency of CO and PO attainment reports inevitably leads to a relatively non-dynamic system. For a dynamic LA system, a real-time overview of CO and PO attainment will be a necessary update to the existing system to locate precisely the divergent outcomes of pedagogic methods, instructional strategies or evaluation strategies. An increase in the frequency of attainment report assessment or real-time monitoring of CO and PO attainment would require significant organisational changes. Amirtharaj et al. (2022) list out a series of organisational structures which allow for reconfigurations to take into account the CO and PO attainments. While at the top of the organisational structure remains the Board of Studies, the lower-level ranges from class monitoring committee meetings and faculty council meetings. A robust LA strategy would require formalizing to an extent and empowering individual teachers to conduct pedagogic experiments and changes to ascertain different strategies. The current organisational structure remains relatively bureaucratic and slow to effect change.

Another important aspect that requires flagging is that the existing documentation exercise for NAAC has been noted to be "too intensive and overwhelming for the HEI and must be rationalized and reduced if possible" (Patwardhan et al., 2022, p. 38). While the documentation exercise and the necessary data entry could be automated through learning management systems if efficiently deployed, one can also foresee an increased workload on teachers if the organisation is unwilling to reskill its existing labour force or hire a new workforce in place of the work that was automated. If real-time data collection of CO and PO attainment becomes another chore in the long list of tasks that teachers have to do, the quality of the data and deployment of such new approaches will remain suspect.

4.3 Limitations of Existing Data and Outcome-based Approaches

As already noted, the frequency of data collection is relatively low currently. Call for additional granular data about learning processes and outcomes can only be effective if organisational changes are implemented to make them more dynamic. The current organisational mechanisms noted by Amirtharaj et al. (2022) are relatively bureaucratic. Division of existing COs and POs through tools such as Bloom's Taxonomy has been noted to be an effective method of curriculum review to identify tasks where students face hurdles (Teater, 2011). Such meta-classification of COs and POs remains a possible approach to the granular mapping of COs and POs.

While granular data ontologies taking into account LOs remains a possible positivist route, it is also important to flag the limitations of such an approach. Constructionist critiques of learning outcome-based education note that "LOs that are expected to be full-ended and predefined" (Havnes & Prøitz, 2016, p. 219) efface unexpected learning, learning from peers and social surroundings. They also note that "some knowledge may be difficult to specify due to their level of abstraction" (Havnes & Prøitz, 2016, p. 208) and warn against the managerialist tendencies of outcome-based approaches.

5. Conclusion

The field of LA has been noted to be lacking in engagement with educational and learning theories, most often resorting to proxies of learning processes to measure learning. At the same time, authors inspired by the field of critical theory have pointed towards fundamental risks of LA, with increased surveillance, racializing drives of artificial intelligence and machine learning while others have noted the production of new social orderings through fundamental methodological choices of LA. A significant amount of work, going forward, needs to be done to flesh out the learning theories which shall inform the data structures for LA deployment. The current CO and PO mapping and the usage of mastery levels reflect a genealogical trace of work by Benjamin Bloom for example.

By beginning our work on LA, from COs and POs i.e., by centring learning outcomes, we wish to address the existing lacunae in the field of LA which has tended to work without a strong focus on education and learning theories. By focusing on existing data collection, we could reconfigure existing methodologies for LA deployment while addressing the limitations and flaws of the existing data methodologies. By centring on learning outcomes and processes, we also hope to reduce the stress on demographic data. Coupled with that, an acknowledgement of the limitations of LOs and LA will allow for differential student trajectories and non-normative learning behaviours to be accounted for.

6. Limitations and Future Work

The SSRs generated for the NAAC also include information about learning management systems (LMS), their usage and deployment. Information concerning LMS usage in the reports across institutions is very inconsistent. While most institutions suggest the usage of LMS, only a few institutions report in detail the range of LMSs and the nature of their usage. A thorough mapping of LMSs across the country would provide us with much-needed insight into the baseline conditions across various HEIs. The nature of the data in HEIs in India, whether they can be represented in xAPI for example (Bakharia, Kitto, Pardo, Gašević, & Dawson, 2016), would also need to be surveyed and understood. While thinking through LA deployment at a trans/inter-institutional scale it is also important to ascertain the technical and policy context of interoperability of learning records. But it is important at the same time to acknowledge the limitations of LMSs and to think of learning analytics in an offline context. Further experiments with outcome-based education and data-based reconfiguration of the existing approaches would also need to be done to ascertain data ontologies for an LA informed by educational and learning theories while taking into consideration the privacy of the human actors.

Acknowledgements

This work is supported by a grant from Vidyashilp University, Bengaluru. We would like to thank Dr Radhika Lobo for pointing us toward investigating the existing data being generated for the NAAC accreditation purpose.

References

Amirtharaj, S., Chandrasekaran, G., Thirumoorthy, K., & Muneeswaran, K. (2022). A Systematic Approach for Assessment of Attainment in Outcome-based Education. *Higher Education for the Future*, *9*(1), 8–29. https://doi.org/10.1177/23476311211017744

Bakharia, A., Kitto, K., Pardo, A., Gašević, D., & Dawson, S. (2016). Recipe for success: Lessons learnt from using xAPI within the connected learning analytics toolkit. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 378–382. Edinburgh, United Kingdom: ACM Press. https://doi.org/10.1145/2883851.2883882

Chen, B., & Fan, Y. (2018). Learning Analytics: Perspectives from Mainland China. In C. P. Lim & V. L. Tinio (Eds.), *Learning analytics for the global south* (pp. 36–43). Quezon City: Foundation for Information Technology Education and Development.

Eradze, M., Väljataga, T., & Laanpere, M. (2014). Observing the Use of e-Textbooks in the Classroom: Towards "Offline" Learning Analytics. In Y. Cao, T. Väljataga, J. K. T. Tang, H. Leung, & M. Laanpere (Eds.), *New Horizons in Web Based Learning* (pp. 254–263). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-13296-9_28

Gašević, D. (2018). Include Us All! Directions for Adoption of Learning Analytics in the Global South. In C. P. Lim & V. L. Tinio (Eds.), *Learning analytics for the global south* (pp. 1–21). Quezon City: Foundation for Information Technology Education and Development. Retrieved from http://dl4d.org/wp-content/uploads/2018/03/Learning-Analytics-Full-Paper-2.pdf

Guzmán-Valenzuela, C., Gómez-González, C., Rojas-Murphy Tagle, A., & Lorca-Vyhmeister, A. (2021). Learning analytics in higher education: A preponderance of analytics but very little learning? *International Journal of Educational Technology in Higher Education*, *18*(1), 23. https://doi.org/10.1186/s41239-021-00258-x

Havnes, A., & Prøitz, T. S. (2016). Why use learning outcomes in higher education? Exploring the grounds for academic resistance and reclaiming the value of unexpected learning. *Educational Assessment, Evaluation and Accountability*, *28*(3), 205–223. https://doi.org/10.1007/s11092-016-9243-z

National Assessment and Accreditation Council. (2018). *NAAC Institutional Accreditation: Manual for Self-Study Report Universities*. Bengaluru: National Assessment and Accreditation Council. Retrieved from National Assessment and Accreditation Council website: http://www.naac.gov.in/images/docs/Manuals/University-Manual-24th-October-2018.pdf

Patwardhan, B., Mohanan, K. P., Mohanan, T., Ramakrishnan, V., Grewal, R. S., Shankar, D., … Sharma, S. C. (2022). *Re-Imagining Assessment and Accreditation in Higher Education in India*. Bengaluru: National Assessment and Accreditation Council. Retrieved from National Assessment and Accreditation Council website: http://www.naac.gov.in/images/docs/notification/Re-Imagining_Assessment_and_Accreditation_in_Higher_Education_in_India_2207202.pdf

Prinsloo, P. (2018). Context Matters: An African Perspective on Institutionalizing Learning Analytics. In C. P. Lim & V. L. Tinio (Eds.), *Learning analytics for the global south* (pp. 24–35). Quezon City: Foundation for Information Technology Education and Development.

Stella, A. (2004). External quality assurance in Indian higher education: Developments of a decade. *Quality in Higher Education*, *10*(2), 115–127. https://doi.org/10.1080/1353832042000230608

Stella, A. (2015). Institutional Accreditation in India. *International Higher Education*, (27). https://doi.org/10.6017/ihe.2002.27.6982

Teater, B. A. (2011). Maximizing Student Learning: A Case Example of Applying Teaching and Learning Theory in Social Work Education. *Social Work Education*, *30*(5), 571–585. https://doi.org/10.1080/02615479.2010.505262