Profiling Engagement in a Virtual Patient Module: A Learning Analytics Approach to Behavioral Data

Lara McBean

Faculty of Arts, University of British Columbia

April 11, 2025

Introduction

In medical education, student engagement is a critical component of learning that directly influences knowledge retention, clinical reasoning, and skill development. As online and blended learning environments have become increasingly common—particularly in response to the COVID-19 pandemic—educators face a growing challenge: ensuring that students engage meaningfully with digital course content outside traditional classroom settings (Aristovnik et al., 2023).

Virtual Patient (VP) modules are especially valuable among the many tools that support online clinical learning. These simulations offer students a safe, flexible environment to practice diagnostic reasoning, clinical communication, and decision-making skills. In contrast to passive video-based instruction, VP modules often require active participation through written or spoken responses, allowing students to rehearse essential tasks in a controlled setting (Cook & Triola, 2009). Their use is particularly relevant in pre-clinical training and residency preparation, where students begin to bridge the gap between theoretical knowledge and practical application.

This study focuses on the Mr. Kato ophthalmology module—a virtual patient case designed to help medical students practice patient history-taking through interactive content and required written responses. Because the module is self-paced and unsupervised, understanding how students engage with it is essential for evaluating its instructional effectiveness. However, assessing engagement in online environments requires behavioural and linguistic indicators to capture meaningful interaction.

To investigate patterns of student engagement, we draw from the behavioural framework of learning analytics, combining time-on-task data with natural language processing (NLP) analysis of student-generated text. By examining how long students spend on each page and the quality of their written responses, we aim to capture richer dimensions of engagement. These data are analyzed using Latent Profile Analysis (LPA), a person-centered statistical technique identifying subgroups of students with similar behaviour patterns. This combined approach offers a novel contribution to the study of online clinical education by integrating behavioural trace data with linguistic indicators of effort. It also enables us to compare how engagement profiles may differ across student cohorts who completed the module at different times.

Understanding how students interact with virtual simulations like Mr. Kato has implications for instructional design, especially in courses that rely on asynchronous, self-paced learning. With this in mind, we ask:

> *What are the differences in high, moderate, and low engagement profiles between two groups of different time frames for medical students in the Mr. Kato ophthalmology module?*

By identifying behavioural engagement profiles and comparing their characteristics across groups, we aim to inform future instructional strategies, support personalized feedback, and ultimately enhance the learning experience in virtual clinical environments.

Background

Theoretical Frameworks of Student Engagement

Student engagement is a multidimensional construct commonly divided into three aforementioned frameworks: behavioural, cognitive, and affective engagement (Fredricks et al., 2004). Behavioural engagement refers to the actions students take in their learning environment, such as participating in activities, staying on task, and demonstrating persistence. Cognitive engagement relates to the investment in learning, involving self-regulation, metacognition, and deep learning strategies. Affective engagement focuses on emotional responses in learning contexts, such as interest, enjoyment, or belonging.

Behavioural engagement is the most directly measurable in online and blended environments through digital trace data, making it a common focus in learning analytics research (Henrie et al., 2015). Berman and Artino (2020) further emphasize that behavioural engagement can be observed through indicators such as time on task, accuracy, and the completion of required tasks. These proxies offer practical ways to assess student involvement in online modules where direct observation is not feasible.

Engagement in Medical Education and Virtual Patients

Virtual patient (VP) simulations are increasingly used in clinical education to train diagnostic reasoning, patient communication, and procedural skills. These systems offer flexible, safe environments for students to engage in complex clinical scenarios before entering high-stakes, real-world environments (Cook & Triola, 2009). Studies have shown that when students actively engage in VP modules—particularly through reflective tasks and written responses—they develop stronger clinical decision-making abilities and more profound understanding (Consorti et al., 2012).

However, student engagement in VP modules can vary widely depending on the design of the module and the learner's motivation. Online environments lack the affordances of physical classrooms, such as immediate instructor feedback and peer interaction, making it harder to detect disengagement without analyzing student behaviour in detail (Bates & Khasawneh, 2007). In this context, using behavioural data, such as how long students spend on different types of content and the quality of their written responses, becomes crucial for identifying engagement patterns.

Behavioural Metrics: Time and Effort

Time-on-task is one of the most common engagement indicators in online learning environments. Research has consistently shown that the amount of time students spend interacting with content is associated with performance and learning outcomes (Jo, Kim, & Yoon, 2015). However, time alone is insufficient, as students may click through content without meaningful interaction. To improve the validity of engagement metrics, researchers often pair time-on-task with other variables such as quiz attempts, mouse movements, or textual inputs.

This study operationalizes effort through natural language analysis of student responses within the VP module. As progression requires a response on all pages prompting it, the depth and tone of those responses offer meaningful insight into student engagement. Short or meaningless answers may indicate minimal engagement, while longer, more thoughtful responses suggest cognitive involvement and effort (Wen et al., 2014). Researchers can better understand behavioural engagement by analyzing response length, complexity, and sentiment through NLP tools.

Natural Language Processing in Learning Analytics

NLP techniques, including sentiment analysis, readability scoring, and semantic richness detection, are increasingly used to quantify textual engagement in educational settings (Jim et al., 2024). Sentiment has been linked to affective and behavioural engagement in discussion forums and open-response assignments, while complexity scores help indicate cognitive effort.

Although sophisticated NLP models can offer profound insights, lightweight approaches—such as polarity scoring and fundamental syntactic analysis—are often helpful in detecting patterns of low-effort versus high-effort responses in constrained environments like virtual simulations (Sultan et al., 2016).

Latent Profile Analysis in Engagement Research

Latent Profile Analysis (LPA) is a person-centered statistical technique used to uncover hidden subgroups (or "profiles") in a dataset based on continuous variables. Unlike traditional variable-centred approaches (e.g., regression), LPA categorizes individuals into latent classes based on their response patterns. It is particularly suitable for analyzing engagement behaviours across different learners (Gilbert et al., 2023).

In educational research, latent constructs have been identified to identify student subtypes based on motivation, strategy use, and engagement metrics (Jovanović et al., 2019). For example, some studies have found distinct profiles of “engaged,” “strategic,” and “disengaged” learners based on time-on-task and self-reported effort. These profiles can help instructors tailor interventions, adapt instructional strategies, and understand how different learners respond to the same materials.

Based on their time and effort data, this study uses LPA to group medical students into high, moderate, and low engagement profiles. By comparing these profiles across two cohorts, we aim to understand how students engage with the Mr. Kato ophthalmology module and how engagement patterns may shift over time or across different independent groups.

Methods

Data Sources

Interaction data for this study was collected from the Mr. Kato ophthalmology module, which is hosted within the Entrada learning management system (LMS) used by the University of British Columbia’s Faculty of Medicine. All module interactions were automatically recorded via the Experience API (xAPI), a standard for learning activity tracking that allows detailed logs of student behaviours. These logs were stored in the platform’s Learning Record Store (LRS) as JSON objects. They provided a comprehensive account of how users navigated the module, including timestamps, page views, response submissions, and activity types.

The dataset includes records from November 1, 2022, to November 1, 2024, covering 404 unique students. All identifying information was removed before analysis. Anonymization included replacing names with person numbers and removing emails. The dataset was cleaned to remove non-student interactions, such as those logged by staff or instructional designers who may have accessed the module for administrative purposes.

Having cohort information would help compare how different groups engage with the content over time. However, the dataset did not have this information available, so to simulate this, the dataset was divided into two independent groups based on time of interaction:

- Group 1: Students who interacted with the module from November 1, 2022, to October 31, 2023
- Group 2: Students who interacted with the module from November 1, 2023, to November 1, 2024

Each student's interactions were aggregated and structured into two main data frames for analysis—one focused on time spent and the other on text-based effort.

Computing Time-on-Task

The first behavioural variable—time—was calculated by summing the total number of seconds each student spent on each module page. We opted for total time per content category instead of other aggregates like mean to minimize distortion caused by students who might skim or click rapidly through slides without engaging with the content.

Each module page was manually categorized into one of four types:

1. Assessment – pages requiring written student responses
2. Informative – pages providing background or case information

3. Interactive – pages for reviewing course material
4. Logistics – pages providing module details and survey

Summing time across categories allowed us to construct a profile for each student. These were used as input variables in the latent profile analysis.

Measuring Effort via Text Responses

The second key behavioural variable—effort—was operationalized using students' written responses within the module. Since progression through the module required completing text boxes when prompted, these responses offered a valuable proxy for student effort and cognitive engagement.

To assess the quality of these responses, we built a scoring system using a combination of rule-based heuristics and NLP tools. Each response was scored from 0 to 10 based on the following components:

- Complexity: Responses with only a single letter or word received low scores, while complete sentences received higher scores.
- Sentiment: We scored responses using the VADER sentiment analyzer. Sentiment extremes (high negativity or sarcasm) were flagged but not penalized unless non-informative.
- Length: Responses are normalized to a maximum of 100 words, as the questions asked do not prompt longer responses.

- Word Count: Provided more data in cases where students exceeded 100 words.

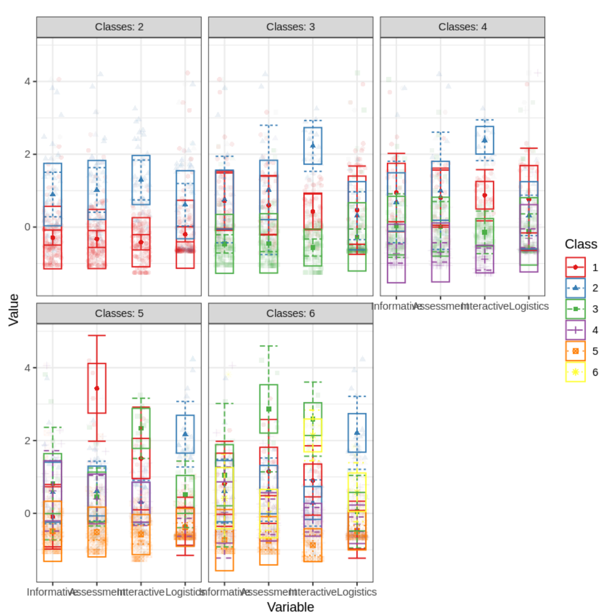
Removing Outliers

Initially, a standard interquartile range (IQR) method was applied, but it proved too rigid and excluded high-effort students. We manually adjusted thresholds to better represent valid values while removing erroneous entries.

Latent Profile Analysis (LPA)

Time Profiles

The time LPA model included four variables per student, each representing the total number of seconds spent in one of the module's content categories (Assessment, Informative, Interactive, and Logistics). We tested models with two to six latent classes, as seen in Figures 1a and 1b.



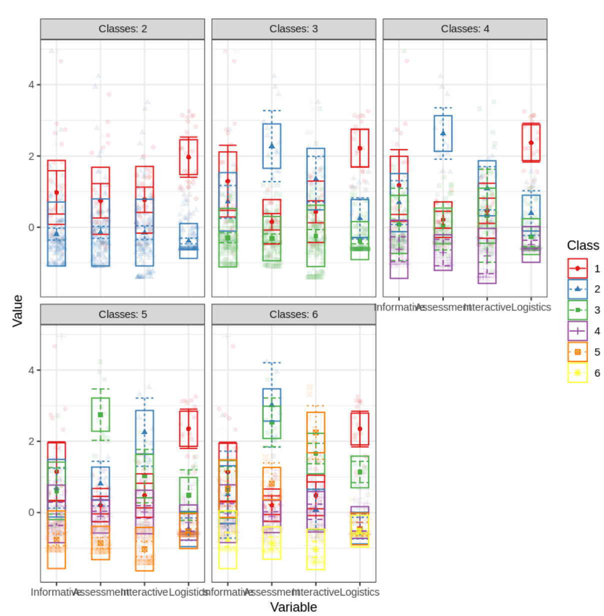


Figure 1a. Group 1: LPA Class Testing

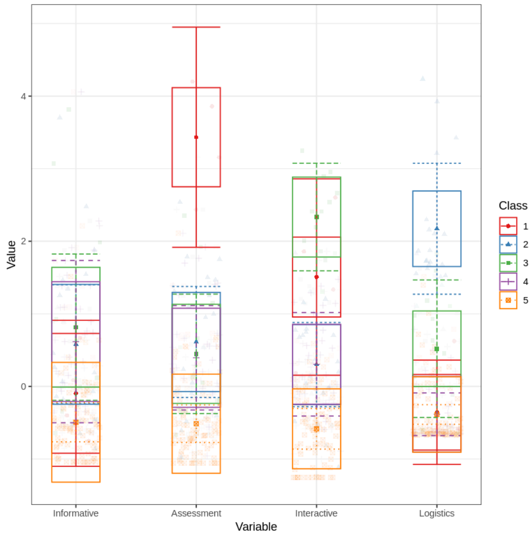
Figure 1b. Group 2: LPA Class Testing

Model selection is based on:

- Bayesian Information Criterion (BIC)
- Akaike Information Criterion (AIC)
- Entropy
- Bootstrapped Likelihood Ratio Test (BLRT) values and p-value

The best model has the lowest BIC and AIC values, or when the ratio between k and k-1 plateaus closer to 1. Entropy values of > 0.80 are preferred, indicating a clear separation of classes. The BLRT compares the fit of a model with k classes to a model with k–1 classes. A high BLRT value with p-values of < 0.05 indicates significant model improvement.

For both groups, the best-fitting model was a five-profile solution, as seen in Tables 1 and 2 of the Appendix. Although a six-profile model offered slightly better AIC/BIC scores in Group 2, one profile contained too few students (<5%) to justify its inclusion.



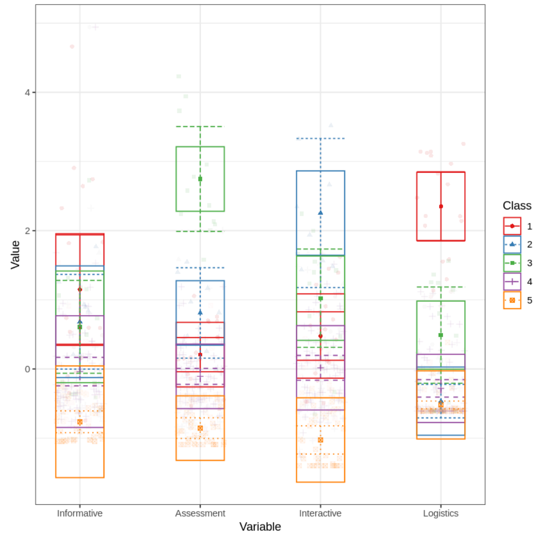


Figure 3a. Group 1: 5-Class Profiles Figure 3b. Group 2: 5-Class Profiles

Effort Profiles

A similar process was followed for the effort data, where the input variables were average response scores across module sections. A three-profile solution best described both groups, with well-separated clusters representing high, moderate, and low text-based effort. See Tables 3 and 4 in the Appendix.

Profile Labeling and Validation

Once the optimal number of profiles was determined for each dataset, we assigned qualitative labels—high, moderate, and low engagement—based on the relative mean scores within each profile. For example, the profile with the highest average total time across all content categories was labelled "High Time Engagement." Similarly, the profile with the longest, most complex responses was labelled "High Effort Engagement."

Results

Overview of Statistical Analysis

We conducted the Mann-Whitney U test to examine whether user engagement profiles (High, Moderate, Low) exhibited significant variation in time spent per page and interaction with textual features. This data is heavily skewed to the right, and this non-parametric test is suitable for comparing two independent groups when the data does not necessarily follow a normal distribution.

Assumptions and Considerations

Before conducting the analysis, the assumptions of the Mann-Whitney U test were evaluated. The first assumption—independent observations—was met, as each user was uniquely associated with a single profile group and did not appear in multiple groups. Second, the engagement profile variable was considered ordinal, as profiles were categorized based on increasing levels of interaction or activity (i.e., Low < Moderate < High). Third, the Mann-Whitney U test assumes an adequate sample size, particularly for reliable statistical significance and effect magnitude inference.

Some concerns arose at this third assumption. Although many comparisons had acceptable sample sizes ($n > 30$ in both groups), specific profile-content groupings—particularly within the High-profile category for time engagement—contained smaller sub-samples ($n1 = 17$ and $n2 = 20$). Such limitations may reduce the power of the analysis and increase the likelihood of Type II errors (i.e., failing to detect an actual effect). As such, interpretation of the results must be made with caution.

To provide a more nuanced understanding, we report p-values and effect sizes (r), which reflect the practical magnitude of differences independent of sample size. Effect size interpretations follow conventional benchmarks: small (0.1–0.3), medium (0.3–0.5), and large (>0.5). Detailed results can be found in Tables 5 and 6 in the Appendix.

Time Spent per Page

The analysis of time spent per page identified several statistically significant differences between corresponding engagement profiles in the two groups. These effects varied by content type and engagement profile, with some content categories showing more apparent differentiation than others.

Logistics Pages.

The Logistics content category showed the most substantial divergence, particularly among Low-profile users, where a large effect size was observed ($W = 2647.5$, $p < .001$, $r = .511$). Differences were also found among Moderate-profile users ($W = 3734.0$, $p = .001$, $r = .256$, small effect), while the High-profile comparison yielded a more negligible, non-significant effect ($W = 219.0$, $p = .137$, $r = .244$). These results suggest that logistical content was especially salient in differentiating lower-engagement users across the two groups.

Assessment Pages.

For Assessment pages, High-profile users showed a medium-sized difference in time spent (W =

92.0, $p = .018$, $r = .388$), as did Low-profile users ($W = 2412.0$, $p < .001$, $r = .395$). The Moderate-profile comparison yielded a smaller but statistically significant effect ($W = 3755.0$, $p = .001$, $r = .260$). These findings indicate that engagement with assessment content varied meaningfully between the groups, particularly among users with high and low engagement.

Interactive Pages.

Differences in time spent on Interactive content were most pronounced among High-profile users ($W = 257.0$, $p = .008$, $r = .433$), followed by the Low-profile group ($W = 2398.0$, $p < .001$, $r = .389$), both showing medium effects. The Moderate group showed a more negligible effect ($W = 3514.0$, $p = .017$, $r = .190$). Although statistically significant across all comparisons, the overall magnitude of these differences was more modest relative to logistics or assessment pages.

Informative Pages.

The Informative content category showed the least differentiation between profiles. Among High-profile users, the effect was small and not statistically significant ($W = 132.0$, $p = .253$, $r = .188$). In contrast, both the Low-profile ($W = 2236.5$, $p = .001$, $r = .308$) and Moderate-profile comparisons ($W = 3794.0$, $p = .001$, $r = .271$) showed small to medium effects. These results suggest that general informational content was approached similarly across groups, particularly by more engaged users.

Text Scores

In contrast to time-based data, analyses of textual characteristics—including complexity score, sentiment score, length score, and word count—yielded uniformly negligible to minor effects across all user engagement profiles.

High-profile Group.

None of the textual characteristics reached statistical significance for the High-profile group. Effect sizes were minimal across the board: complexity score ($r = .006$), length score ($r = .085$), sentiment score ($r = .014$), and word count ($r = .087$). All fall within the negligible effect sizes, suggesting that text complexity or sentiment did not drive differential behaviour among highly engaged users.

Low-profile Group.

Results for the Low-profile group showed slightly more variability across textual features; however, none of the differences reached statistical significance. Both complexity score ($r \approx .107$, $p = .135$) and length score ($r \approx .103$, $p = .148$) yielded small effect sizes, while sentiment score ($r \approx .097$, $p = .176$) and word count ($r \approx .028$, $p = .691$) showed negligible differences.

Moderate-profile Group.

For the Moderate-profile group, differences in interaction based on textual features were minimal. Both complexity score ($W = 440.0$, $p = .763$, $r = .039$) and length score ($W = 386.5$, $p = .610$, $r = .066$) yielded negligible effects, indicating virtually no difference in engagement with content of varying complexity or length. Slightly larger, though still small, effects were observed for sentiment score ($W = 489.0$, $p = .290$, $r = .138$) and word count ($W = 357.0$, $p = .333$, $r = .126$).

These findings suggest that users across all engagement types interacted relatively uniformly with text content.

Discussion

The engagement profiles identified in this study—High, Moderate, and Low—offer valuable insights into how students interact with online course materials, particularly when examined through the dual lenses of time spent per page and textual content engagement. These profiles reveal meaningful patterns that reflect behavioural differences and potentially different orientations to self-regulated learning, digital literacy, and motivation. In this discussion, we interpret these profiles, link them to existing literature on engagement in online education, and consider the implications for course design and instructional strategies.

Interpreting the Profiles: Behavioural Patterns and Engagement Types

The most substantial differences in time-on-page for students in the High-engagement profile were observed for Assessment and Interactive content, both showing medium effects. However, these effects moved in opposite directions across the two groups. For Assessment, Group 2 spent more time than Group 1, suggesting a stronger emphasis on reflection or evaluative tasks. Conversely, Group 1 exhibited a higher average time-on-page for Interactive and Logistics content. This may reflect a preference for active or navigational content among those users, possibly indicating different styles of strategic engagement. Notably, the Assessment subgroup included fewer than 30 participants per group, which may limit statistical power and inflate the risk of Type II error. Informative content also showed slightly higher means for Group 2, but the effect was small and non-significant. Altogether, these patterns suggest that highly engaged users in both groups focused time on cognitively demanding tasks but differed in their areas of emphasis, possibly due to differing task strategies or motivational orientations.

However, textual analysis of their written responses showed negligible differences across all linguistic features (sentiment, complexity, length, and word count). This indicates that, although behavioural engagement (time spent) diverged somewhat between groups, their writing performance remained stable, reinforcing the interpretation that these learners are consistent in output, regardless of contextual variation (Winne & Hadwin, 2008).

The Moderate-engagement group showed small but statistically significant differences across all content categories, with Group 1 spending more time than Group 2 in each case. Although these effects were modest, the consistent trend may suggest that later cohorts exhibited lower levels of behavioural engagement. While the current data do not allow for causal conclusions, the pattern may reflect a potential decline in engagement over time, possibly due to reduced novelty, changing external demands, or diminished instructional presence in later course iterations. Textual effort scores revealed minor differences in sentiment and word count, with Group 1 again showing slightly more expressive or extensive responses. Though these effects were not statistically significant, they are directionally consistent with the time-based findings and may indicate mild motivational or attentional differences between cohorts. Overall, this profile suggests a functional but surface-level engagement pattern, and the observed decrease in time and output among later users raises questions about how to sustain engagement across repeated offerings or over the academic term.

In the Low-engagement profile, Group 1 consistently spent more time than Group 2 across all content types, with extreme differences for Logistics pages, where a large effect was observed ($r = .511$). This across-the-board pattern points toward a possible cohort-level decline in engagement, where later users may be skimming or bypassing content more frequently. While increased efficiency cannot be ruled out, the absence of any corresponding improvement in textual effort—where differences were negligible to small and not statistically significant—suggests this was not a result of greater fluency or confidence.

Instead, the combination of lower time investment and minimal expressive output in Group 2 supports the interpretation that engagement may have eroded over time, at least for this lower-profile group. These learners may have approached the module with less motivation, less support, or fewer perceived stakes. While we cannot confirm this pattern without longitudinal tracking, the findings highlight a potential need for enhanced scaffolding or motivational interventions in later course cohorts to prevent disengagement (Fredricks et al., 2004).

Connection to Literature: Multidimensional Engagement

The present findings reinforce the multidimensional nature of engagement—spanning behavioural, cognitive, and emotional dimensions—as highlighted in prior research (Fredricks et al., 2004; Sinatra et al., 2015). Among the two metrics examined, time spent per page, as a proxy for behavioural engagement, provided clearer and more consistent differentiation across engagement profiles and cohorts than textual effort scores. In contrast, measures like text complexity, sentiment, and word count produced negligible to small effects with limited statistical significance, particularly in the High-profile group.

This supports existing cautions in the learning analytics literature about over-relying on linguistic features as proxies for engagement (Gašević et al., 2016). While these textual characteristics may reflect student cognition or affect aspects, they did not meaningfully distinguish between cohorts or profiles in this context. This suggests that observable behaviours—such as navigation and time allocation—remain more reliable engagement indicators in self-paced, asynchronous learning environments.

Notably, the most potent effects emerged for logistics and assessment pages, particularly in the Low-profile group, pointing to the salience of course structure and evaluative components in shaping student behaviour (Kizilcec et al., 2013). These content types are typically tied to clarity of expectations and performance outcomes, making them natural focal points for students navigating a flexible online module.

Finally, the consistent finding that earlier cohorts (Group 1) spent more time on content across both the Low and Moderate profiles may hint at a cohort-level decline in engagement over time. While causality cannot be established, and individual-level effort trajectories remain untracked, this pattern raises important questions about sustaining motivation and attention in recurring course offerings. It suggests a need to adapt design or support structures to maintain engagement as the novelty of the experience fades or external conditions shift across iterations.

Limitations

Several methodological limitations must be acknowledged:

1. Outlier removal sensitivity: Despite adjustments to the IQR method, legitimate data points from highly engaged or disengaged students may have been excluded, potentially biasing the results.
2. Assumed first-time access: The dataset treats the first recorded interaction during the study period as the student's first time engaging with the module. Students, particularly

those in Group 1, may have accessed or reviewed the module before data collection. This implies that the data captures review behaviours, not initial engagement.

3. Behavioural focus: The study is limited to the behavioural dimension of engagement and does not account for affective (e.g., emotional investment) or cognitive (e.g., metacognition, deep learning strategies) dimensions, which are also integral to comprehensive engagement frameworks.
4. Text scoring subjectivity: While NLP tools assessed sentiment and some complexity indicators, evaluating student responses still involved manual components. This introduces potential bias and inconsistency in scoring.
5. Module structure constraints: The module's linear structure may have constrained natural variations in engagement. For example, students were required to complete specific tasks to move forward, which may have masked differences in voluntary effort.
6. Engagement Profile Structure: Profiles were treated as independent and ordinal, reflecting their conceptualization as distinct levels of engagement (Low, Moderate, High) rather than continuous measures. However, future work could explore more nuanced clustering approaches (e.g., latent class analysis or k-means clustering) to validate these groupings and refine comparisons.

Future Applications

This methodology offers promising avenues for future research and application. In particular, the LPA framework can be integrated with learning trajectory models such as VaSSTra (Variable-State Student Trajectory Analytics) to examine how engagement profiles evolve across multiple modules or in the curriculum.

Moreover, this behavioural engagement framework can inform instructional design by identifying pain points—segments where students consistently exhibit lower engagement. Such insights can guide iterative redesigns of digital learning experiences to promote deeper, more sustained engagement.

Ultimately, by leveraging these data-driven methods, medical educators can better align online modules like Mr. Kato's with the realities of clinical education, narrowing the gap between digital and face-to-face learning environments.

Conclusion

This study examined how medical students with varying behavioural engagement profiles—High, Moderate, and Low—interacted with a self-paced virtual patient module in an online ophthalmology course. By combining learning analytics with NLP and non-parametric statistical methods, we identified distinct patterns of engagement that varied more strongly by time-on-task than by textual content features. Integrating this approach with longitudinal trajectory models could deepen our understanding of how engagement evolves over time and across domains. Ultimately, aligning digital modules with the realities of clinical learning—through better design, feedback, and scaffolding—can help ensure that online experiences inform and meaningfully prepare students for the demands of residency and practice.

High-engagement users exhibited relatively consistent behavioural patterns, with differences in time spent across content types reflecting a strategic orientation. Notably, users in the second cohort with high engagement profiles spent more time on Assessment and Informative content. In comparison, the earlier cohort spent more time on Interactive and Logistics pages—suggesting potential variation in task focus or course navigation strategies. However, these differences should be interpreted cautiously given the small subgroup sizes, particularly for the High profile.

Moderate and Low-engagement users in the second cohort consistently spent less time across all content types than their earlier counterparts. While this may indicate a general decline in engagement over time, the independent nature of the cohorts prevents firm conclusions. Still, the pattern raises important questions about how later cohorts interact with self-paced modules and whether course fatigue, shifting expectations, or contextual factors may be at play.

Across all profiles, linguistic features of student responses—such as complexity, sentiment, and word count—exhibited negligible to minor effects, reinforcing that static textual characteristics are limited engagement indicators. This supports prior research cautioning against overreliance on text analytics alone and highlights the continued relevance of behavioural metrics like time-on-task in digital learning environments.

These findings affirm the multidimensional nature of engagement and suggest several actionable directions for design: streamlining logistical content to minimize confusion, leveraging assessment activities to promote reflection, and proactively supporting users who exhibit signs of early disengagement. While limited by sample size and its behavioural focus, this study offers a practical, replicable approach to profiling learner engagement and identifying points of pedagogical intervention in self-paced online modules.

References

Aristovnik, A., Karampelas, K., Umek, L., & Ravšelj, D. (2023). Impact of the COVID-19 pandemic on online learning in higher education: A bibliometric analysis. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1225834

Bates, R., & Khasawneh, S. (2007). Self-efficacy and college students' perceptions and use of online learning systems. *Computers in Human Behavior*, *23*(1), 175–191. https://doi.org/10.1016/j.chb.2004.04.004

Berman, N. B., & Artino, A. R. (2018). Development and initial validation of an online engagement metric using virtual patients. *BMC Medical Education*, *18*(1), 213. https://doi.org/10.1186/s12909-018-1322-z

Consorti, F., Mancuso, R., Nocioni, M., & Piccolo, A. (2012). Efficacy of virtual patients in medical education: A meta-analysis of randomized studies. *Computers & Education*, *59*(3), 1001–1008. https://doi.org/10.1016/j.compedu.2012.04.017

Cook, D. A., & Triola, M. M. (2009). Virtual patients: A critical literature review and proposed next steps. *Medical Education*, *43*(4), 303–311. https://doi.org/10.1111/j.1365-2923.2008.03286.x

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gasevic, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Gilbert, W., Bureau, J., Diallo, A., Guay, F., & Morin, A. (2023). *Implications and considerations in using latent profile and transition analyses in longitudinal research within Self-determination theory*.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, *6*, 100059. https://doi.org/10.1016/j.nlp.2024.100059

Jo, I.-H., Kim, D., & Yoon, M. (2015). Constructing Proxy Variables to Measure Adult Learners' Time Management Strategies in LMS. *Journal of Educational Technology & Society*, *18*(3), 214–225.

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*, 74–85. https://doi.org/10.1016/j.iheduc.2017.02.001

Kassab, S. E., El-Sayed, W., & Hamdy, H. (2022). Student engagement in undergraduate medical education: A scoping review. *Medical Education*, *56*(7), 703–715. https://doi.org/10.1111/medu.14799

Kizilcec, R., Piech, C., & Schneider, E. (2013). Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. In *ACM International Conference Proceeding Series* (p. 179). https://doi.org/10.1145/2460296.2460330

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The Challenges of Defining and Measuring Student Engagement in Science. *Educational Psychologist*, *50*(1), 1–13. https://doi.org/10.1080/00461520.2014.1002924

Sultan, M. A., Salazar, C., & Sumner, T. (2016). Fast and Easy Short Answer Grading with High Accuracy. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1070–1075). Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-1123

Wen, M., Yang, D., & Rosé, C. (2014). *Sentiment analysis in MOOC discussion forums: What does it tell us?*

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). Lawrence Erlbaum Associates Publishers.

Appendix

Table 1. Group 1: LPA values for Time

Class	AIC	BIC	Entropy	Prob_min	Prob_max	n_min	n_max	BLRT (val)	BLRT (p)
2	1719.16	1759.14	0.831	0.907	0.969	0.244	0.756	119.07	0.01
3	1693.58	1748.94	0.803	0.878	0.930	0.094	0.600	35.58	0.01
4	1682.71	1753.44	0.778	0.844	0.978	0.088	0.450	20.87	0.01
5	1601.13	1687.24	0.854	0.765	0.992	0.025	0.569	91.58	0.01
6	1599.40	1700.88	0.823	0.783	0.976	0.031	0.450	11.74	0.198

Table 2. Group 2: LPA values for Time

Class	AIC	BIC	Entropy	Prob_min	Prob_max	n_min	n_max	BLRT (val)	BLRT (p)
2	1660.96	1700.86	0.957	0.964	0.993	0.164	0.836	165.92	0.01
3	1581.51	1636.75	0.957	0.899	0.990	0.094	0.786	89.45	0.01
4	1549.59	1620.17	0.802	0.823	0.979	0.082	0.472	41.92	0.01
5	1509.18	1595.11	0.875	0.833	0.998	0.050	0.535	50.41	0.01
6	1479.16	1580.44	0.898	0.841	1.000	0.031	0.535	40.01	0.01

Table 3. Group 1: LPA values for Text Scores

Class	AIC	BIC	Entropy	Prob_min	Prob_max	n_min	n_max	BLRT (val)	BLRT (p)
2	1524.86	1567.21	0.974	0.986	0.996	0.302	0.698	676.62	0.01
3	1163.92	1222.56	0.972	0.982	0.992	0.130	0.568	370.93	0.01
4	965.83	1040.76	0.966	0.929	0.998	0.109	0.521	208.09	0.01
5	920.84	1012.05	0.944	0.908	0.997	0.109	0.401	54.99	0.01

Table 4. Group 2: LPA values for Text Scores

Class	AIC	BIC	Entropy	Prob_min	Prob_max	n_min	n_max	BLRT (val)	BLRT (p)
2	1464.23	1506.85	0.980	0.994	0.995	0.413	0.587	782.65	0.01
3	1085.93	1144.94	0.979	0.983	0.997	0.199	0.541	388.30	0.01
4	947.60	1023.00	0.971	0.943	0.995	0.122	0.490	148.33	0.01
5	957.60	1049.39	0.777	0.000	0.995	0.000	0.485	0.00	1.00

Table 5. Time: Mann-Whitney U-Test Results

Variable	Profile	W	p-val	Effect Size	Group 1 (n)	Group 2 (n)	Interpretation
Assessment	High	92.0	0.018	0.388	17	20	Medium effect
Informative	High	132.0	0.253	0.188	17	20	Small effect
Interactive	High	257.0	0.008	0.433	17	20	Medium effect
Logistics	High	219.0	0.137	0.244	17	20	Small effect
Assessment	Low	2412.0	0.000	0.395	87	37	Medium effect
Informative	Low	2236.5	0.001	0.308	87	37	Medium effect
Interactive	Low	2398.0	0.000	0.389	87	37	Medium effect
Logistics	Low	2647.5	0.000	0.511	87	37	Large effect
Assessment	Moderate	3755.0	0.001	0.260	56	102	Small effect
Informative	Moderate	3794.0	0.001	0.271	56	102	Small effect
Interactive	Moderate	3514.0	0.017	0.190	56	102	Small effect
Logistics	Moderate	3734.0	0.001	0.256	56	102	Small effect

Table 6. Text Scores: Mann-Whitney U-Test Results

Variable	Profile	W	p-val	Effect Size	Group 1 (n)	Group 2 (n)	Interpretation
Complexity Score	High	2151.0	0.947	0.006	57	76	Negligible effect
Length Score	High	1950.0	0.327	0.085	57	76	Negligible effect
Sentiment Score	High	2203.0	0.868	0.014	57	76	Negligible effect
Word Count	High	1946.0	0.318	0.087	57	76	Negligible effect
Complexity Score	Low	5310.5	0.135	0.107	100	96	Small effect
Length Score	Low	5294.5	0.148	0.103	100	96	Small effect
Sentiment Score	Low	5261.5	0.176	0.097	100	96	Negligible effect
Word Count	Low	4648.0	0.691	0.028	100	96	Negligible effect
Complexity Score	Moderate	440.0	0.763	0.039	35	24	Negligible effect
Length Score	Moderate	386.5	0.610	0.066	35	24	Negligible effect
Sentiment Score	Moderate	489.0	0.290	0.138	35	24	Small effect
Word Count	Moderate	357.0	0.333	0.126	35	24	Small effect