Research Article

Biometric Key Logging in Authentic Learning Environments: Introducing the 'Cursive' Standard and Collection Tools

Joeseph Thibault, *Cursive Technology*

Anissa Sorokin

Structured Abstract

- **Identification of Innovation:** The writing process can be quantified by the key strikes created by a writer using a standard keyboard, also known as biometric key log data. Compared to static text (completed essays), which can be easily collected in various file formats, key log data must be collected at the time of the writing using specific technological resources and cannot be extracted from static documents. When key log data has been collected from participants, the datasets reflect contrived settings and unnatural learning experiences. These nuances have likely hindered and slowed research of biometric key log data in education. This paper introduces and advocates for a standard format of key log data for interoperability in research and practice and introduces two collection tools: an open-source plugin that turns the learning management system (LMS) into a key log data capture tool and a browser extension that does the same.
- **Exposition of Innovation:** In this paper, several biometric key log data collection approaches are introduced–a plugin to the Open Source Moodle project and a browser extension for Chrome–alongside a potential framework to standardize data collection using Javascript Object Notation (JSON). The collection tools aim to provide institutions, researchers, and even individuals with the ability to collect

data more widely while maintaining controls over their data to ensure privacy and security; the framework for standardization is designed to create a common and easy-to-implement dataset. This could help to grow a shareable dataset, with the potential to advance biometric typing research and ensure that institutions retain their ability to move between machine learning models, tools, and commercial services all leveraging a common standard and preventing "vendor lock."

- **Application of Innovation:** Typing is a uniquely human activity that relays thoughts through manually manipulating a keyboard. This survey includes a history of key log data, common and nascent biometric typing uses, and growing interest in research related to quality writing and cognition. A standard dataset ensures the easy application of new machine learning models, feature extraction, and comparison between labeled datasets to accelerate research (and the application of research in practical settings). By showcasing and sharing open source and freely available collection tools already leveraging the standard, we envision a near future where this data is safely collected, shared, analyzed, and benefits presented to teachers and learners in meaningful and actionable ways.
- **Directions for Further Research:** The data schema represents the minimum foundational data required for biometric key log research; however, additional contextual data and events can be added to provide greater granularity and additional insights into writing. Biometric key log data offers a rich data source enabling research in multiple fields: cognition, writing quality, health, identification, and feedback. By enabling easier methods of data capture, the authors hope that research on various fronts can be accelerated and new tools can be created to support writers and improve writing pedagogy.

Identification of Innovation (Background)

Writing has long held significance in and outside of academia as a foundational communication skill (Conijn, Roeser, & van Zaanen, 2019), providing a common basis for peer-to-peer and interpersonal communication, inter- and intra-organizational communication and understanding, ultimately enabling single voices to communicate with the world through news and reporting, books and journals, media and entertainment.

Within academia, it's widely used as a means of both formative and summative assessment. It is used as the capstone for college graduation in the form of a final essay, thesis, or dissertation. Writing, in all formats, has been characterized as the closest approximation of thinking. "Writing is not merely a mode of communication. It's a process that, if we move beyond simple formulas, forces us to reflect, think, analyze and reason. The goal of a writing assignment worth its salt is not simply to describe or persuade or summarize: it's to drive students to make sense of difficult material and develop their own distinctive take" (Mintz, S., 2021)

While the technologies designed to support human-produced writing (e.g., the typewriter, spellcheckers, etc.) have historically generated concern from those who have seen such tools as facilitating the erosion of necessary skills, more recent technology has enabled an acceleration of threats to human-produced writing. For example, advances in Natural Language Processing (NLP), Machine Learning (ML), and Artificial Intelligence (AI) have led to groundbreaking tools such as ChatGPT, Gemini, Grok, Claude, and other generative AI tools that are capable of human-like written communication on a broad spectrum of topics even faster than postulated (Oenbring, 2022). Despite hiccups and criticisms from many angles, these tools, in a short time, have proven to be increasingly fluent in creating text and difficult to discern from human writing (Casal & Kessler, 2023).

Alongside the rise of generative AI tools, a smaller but rapidly growing field of writing analytics has emerged. Initially, the field focused on the quantification of data related to static written text to improve automated feedback and automated scoring systems (Baffour et al., 2022). A growing research body has focused on the rich source of information created through the writing process itself (Leijten & Van Waes, 2013). Data from the writing process can be collected in many forms, from manual observation, the innocuous "Track Changes" of Microsoft Word for editing, to more powerful revision histories in modern web-based Word processors like Google Docs. Even more granular event logging systems can capture events from keyboards, touch pads, touch screens, and mouse use. This granular data can assist researchers in their ability to reconstruct a text or extract various features for analysis veering into the field of biometrics.

Key log programs offer a promising avenue for writing process data collection. As Allen et al. (2016) observed, "...prior writing research ha[d] focused primarily on students' finished writing products and not their moment-by-moment writing processes." The most popular key log programs for collecting moment-by-moment data have included Scriptlog, Translog, and Inputlog (Leijten & Van Waes 2013). Additional options for researchers and practitioners (and ne'er-do-wells) have proliferated through the web, including robust open documentation and shared code, and are available to anyone with a web browser and search bar.

Despite the promise that biometric typing data has presented in recent decades, including research aimed at understanding student writing process and improving student writing process, a major limitation of has been the availability of quality datasets and the nature of the data collection process. The majority of such studies collect writing "within one sitting" (Teh et al., 2013), which impedes researchers' ability to both collect data in an authentic environment and better understand how a person's writing process may change over time. Additionally, data formats have been inconsistent, not always shared, and incompatible with other research initiatives.

The innovation described below attempts to remove these challenges. “Cursive” promotes a standard schema for key log data, including room for contextual information such as grade level, genre, and prompt, and provides extensible tools for collection, including a free key log plugin for the open-source Moodle learning management system (LMS). A stand-alone browser extension for Chrome, which can be enabled by the end-user on word processing sites including Google Docs is also introduced. Both tools capture data in the same format.

The innovation is less a “tool” than a call to recognize and rally around a long-available and valuable standard inherent in biometric typing data. By implementing the tools in the learning management systems or as part of student workflows (through an extension), we can add automatic key log capture functionality into existing educational settings, including long-form essays and research, to facilitate the study of real-world student writing in the natural contexts they are intended to occur. This can be done for single writing sessions, for long form essays written over many sessions, and comprehensive longitudinal studies of writing across courses, semesters, or even years.

Educational research of key log data, historically, has been captured in closed, sterile, and controlled settings during a limited time, potentially creating an observer effect on the data rather than capturing the natural writing practice. Writing, and subsequently biometric typing data, need not be event-based. Instead, it can be recognized and studied as a practice.

The History of Keystroke Data as it Pertains to Writing

It is suggested that the first “key-based” behavioral biometrics were those of telegraph operators and their unique, recognizable rhythms of dots and dashes. The telegraph is just an early input device (Teh et al., 2013) connecting two people. In the mid-20th century, technologists dreamed of new types of devices, not yet called computers. Vannevar Bush’s “As We May Think” described his vision of a “memex” with “direct entry,” which fueled the drive towards more advanced typewriters, word processors, and ultimately computers with keyboards (Bush, 1947).

Today, keyboards represent a ubiquitous standard for input to computers and the creation of written text. Those early echos of behavioral biometrics (recognizing an operator by their dots and dashes) have led to a new and fascinating research landscape for how computers impact and enable the writing process (Selfe & Wahlstrom, 1988) and how biometric key log data captured via direct entry via your standard QWERTY keyboard.

In the last 40 years, key log data has enabled significant research highlighting its usefulness in user identification and verification (Roy et al., 2022), automated essay scoring (Almond et al., 2012), and explorations of task orientation and cognitive load of writers (Chenoweth & Hayes, 2003; Conijn, Roeser, & van Zaanen, 2019; Hayes & Chenoweth, 2007). These explorations and research areas have leveraged statistical analysis of the datasets seeking to understand the minute differences in the extracted features (calculated values) of the data, and most recently, have been subjected to various methods of machine learning to further expose the nuances of each feature within the datasets (Teh et al., 2013).

Unfortunately, key log data is not without controversy. While valuable in research, in the public's eye, it has been lumped in with malicious tactics for identity fraud (Roy et al., 2022) and labeled by some as invasive (Mozur et al., 2022; Stross, 2010). Privacy and security will always be concerns in the use of or capture of key log data; however, responsible capture, anonymization, and sharing can and have led to many insights into teaching, learning, and cognition (Conijn et al., 2022; Leijten, & Van Waes, 2013; Talebinamvar & Zarrabi, 2022).

The value of this data in elevating and improving educational outcomes cannot be understated. However, researchers and educators interested in deploying key log devices and tools must evaluate the ethical, legal, privacy, and transparency considerations with their organizations and participants in advance. It is the view of the authors that the benefits of deployment in specific, authenticated educational settings with student consent far outweigh the concerns and that building solutions that meet local and regional privacy and security requirements could rapidly accelerate adoption and research by proactively solving for those barriers and limitations.

Exposition of Innovation

"Cursive" has developed two keylogging tools and leverages a single standard typing data stream format: a native integration into Moodle LMS and a Chrome-based browser extension directly managed by writers.

The selection of Moodle was partly predicated on the substantial user base already utilizing Moodle as a learning management system, with 100s of millions of users worldwide (Moodle, n.d.). However, four characteristics of Moodle make it stand out as an initial host for the plugin:

1. Open source: the open-source nature of Moodle lends it to customization and a robust community of developers capable of working within the plugin framework. There are no permissions to seek for developing and extending the site, in addition to no fees (except development costs) to register a plugin for the community. The existing install base and the ease and low-cost nature of deploying new Moodle installations for research or learning purposes create significant opportunities for researchers above and beyond existing key log capture tools.
2. Learning Tools Interoperability Provider: Learning Tools Interoperability (LTI) is an international standard for connectivity between learning tools and learning management systems managed by 1Edtech (1Edtech, 2024). Moodle architecture provides site administrators to enable a course or complete site to serve as a tool provider. As such, by building in Moodle, the Cursive plugin can be integrated into any other LMS supporting the LTI standard (this includes D2L, Canvas by Instructure, Blackboard, and others).
3. Vibrant user community: Moodle.org, a site built on the LMS software itself, is a vibrant community of practitioners worldwide seeking and providing assistance daily. The community compounds the value of its open-source nature by providing on-demand, free

support for site administrators and an available forum for discussing research, privacy, security, and pedagogy.

4. Moodle’s default text editor is TinyMCE: as of the most recent versions of Moodle, the default text editor for installations is TinyMCE (MoodleDocs, 2023), a free-to-use text editor used by 100s of 1000s of websites in addition to Moodle’s world-wide installation base. TinyMCE is a versatile web-based writing tool with significant documentation related to JavaScript-based listening events, which can be invoked and extended with limited development (TinyMCE, n.d.).

The second tool is a browser extension for Chrome with various features designed to give the writer control over data collection and summary data availability to third parties. The tool has several features comparable to those described above:

1. Data Collection Controls: a user must opt-in to data collection on any site (this currently has been tested on Google Docs, AI writing sites, OpenAI’s ChatGPT prompt window, and others. Adding compatibility with any online text editor is possible.
2. Data Download and Extraction: only the writer has direct access to download or view captured raw data.
3. Summary Statistics: the data collection on permitted sites is logged automatically and a summary of data is provided to the writer through the extension interface. Summary statistics for a single document can be generated and shared through an unlisted link.

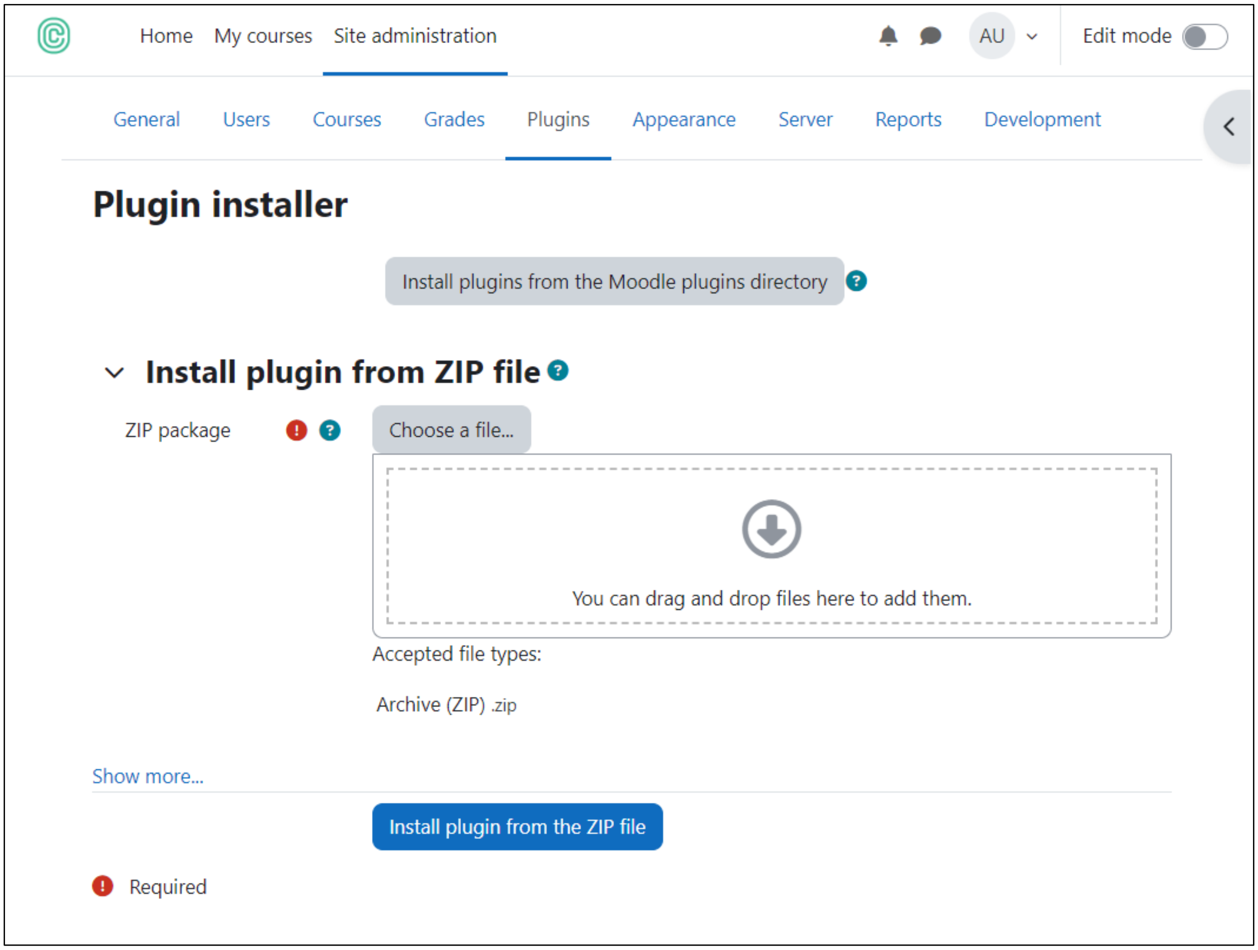
Chrome browser and Google Docs were selected for similar reasons as Moodle: a large userbase, freely available tools (a free personal Google account provides free access to productivity tools including Google Docs), and a robust developer and support community.

Installing the Plugin for Use on an Existing Site

The plugin can be downloaded from a public GitHub repository at https://github.com/cursiveinc/cursive-moodle-tinymce.

The plugin installation can be completed using Moodle’s easy-to-use plugin installation feature via the administrative menu:

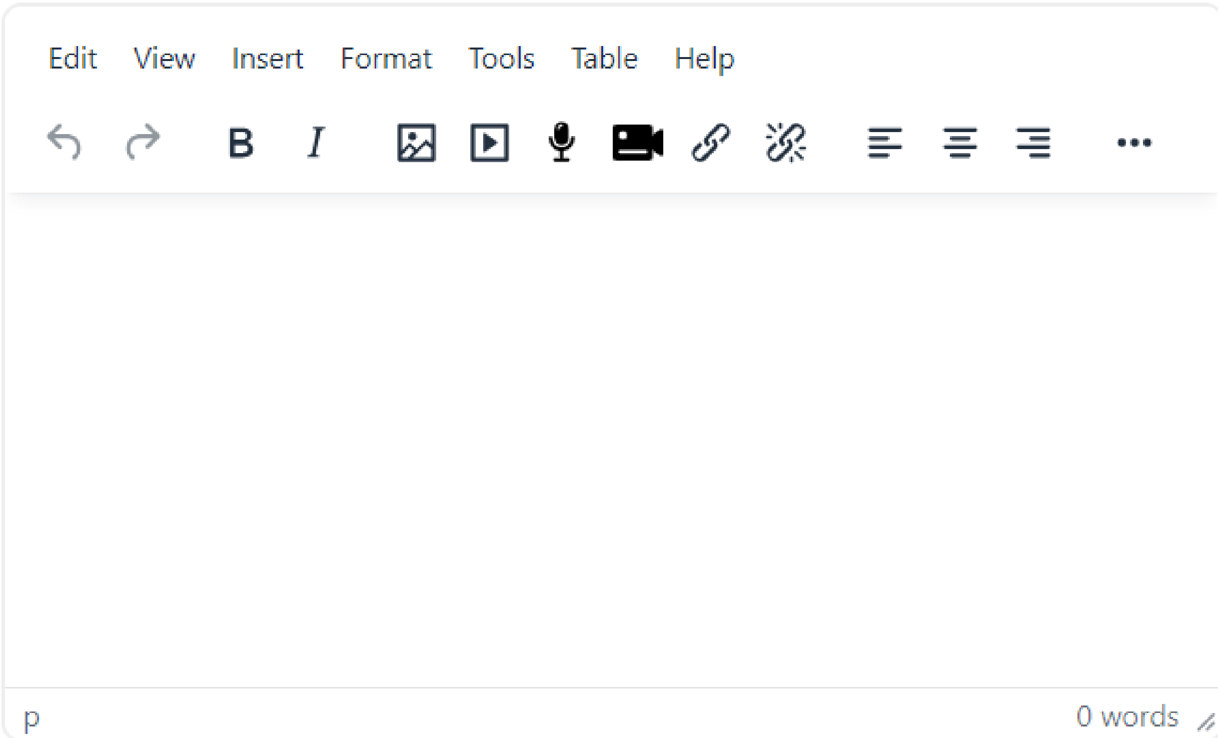
Figure 1



Once installed, the plugin runs for all sessions (student role only) that load the TinyMCE text editor in the following Moodle activities: Quiz (Essay question type), Assignments (Online Text), and Discussion Forums.

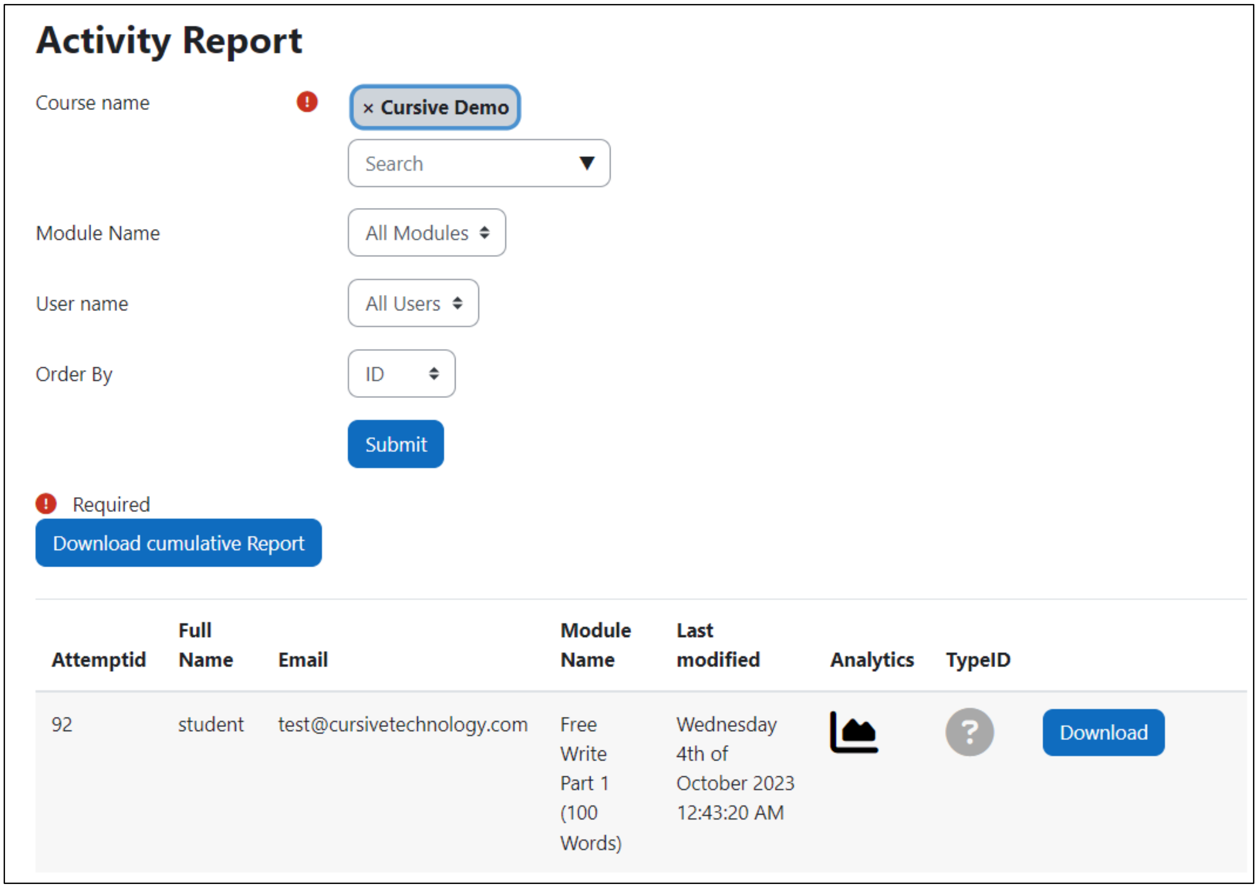
There is no learning curve for students to use the Cursive-enabled interface, the student need only to complete their writing activities within the learning management system:

Figure 2



All completed session files are stored for easy retrieval and download through a distinct URL available to specific roles (those with "teacher" or "administrator" roles) on the site to ensure the privacy and security of files and prevent any unauthorized access (in effort to promote transparency, the plugin also allows students to access their data securely). The report page can be filtered by course, assignment, and student. Each file has an easy-to-access "Download" button for each entry.

Figure 3



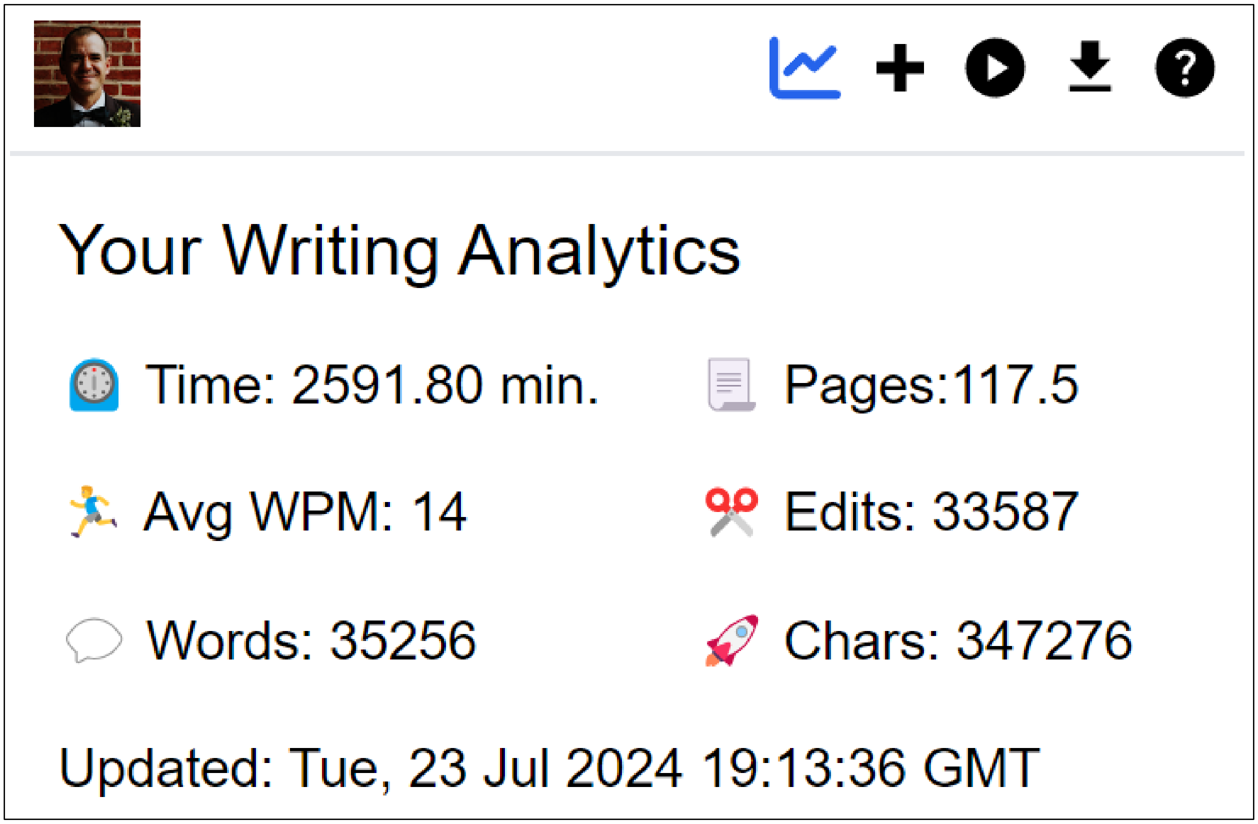
Future iterations of this plugin will add GDPR-compliant controls to automatically request removal of data and to ensure proper controls for both the data controller and data processor while maintaining end-user privacy, security, and consent.

Installing the Browser Extension as a Writer

As a writer, the browser extension can be added through the Chrome web store at https://chromewebstore.google.com/detail/cursive/ejoigjckaakmofdhbnmegfkillcnjgbl. The install and configuration process involves downloading, logging in, and then clicking “add URL” to get started, minimizing user error.

End-user controls include options to manage all aspects of the writing including summary statistics as captured via the writing process in user-allowed web-based text editors.

Figure 4



Data Schema

The data schema is a JavaScript Object Notation or "JSON" (Amazon Web Services, n.d.), which can be manipulated via code (such as Python scripts) or Excel to parse data and run quantitative calculations. The events currently captured are simply keyDown and keyUp, two of the most common and basic attributes in biometric key log data studies (Acien et al., 2022; Allen et al., 2016; Almond et al., 2012; Conijn, Roeser, & van Zaanen, 2019; Conijn et al., 2022; Leijten & Van Waes, 2013; Pisani & Lorena, 2013; Teh et al., 2013). Each event is captured with the following contextual attributes (an example data point is provided for each):

- "resourceId": 92,
- "key": "v",
- "keyCode": 86,
- "event": "keyDown",
- "courseId": "10",
- "unixTimestamp": "1696623514496",
- "clientId": "2df2e6fc-dac2-4706-ac1b-992fb3019343",
- "personId": "57"

It should be noted that the scheme does not include a personal identifier such as name or email, instead the format allows for anonymity based on a unique identifier from the origination. This is intentional as a safeguard for privacy and security to help simplify the sharing of collected datasets for research purposes. The Moodle LMS plugin and Chrome browser extension capture the same format data.

Event attributes such as clientID, resourceID, and courseID may all be used to help to label information for further research, it is also possible to extend context fields without impacting compatibility or interoperability.

Application of Innovation

As described above, the innovation in code is an easy-to-implement, learning management system-compatible key log data capture tool that can be used to facilitate collection of data outside of clinical or prearranged settings. For researchers, it can be implemented in lieu of other commercial or free key log tools and implemented as a single activity, across a course, or complete site providing a highly customizable learning platform and environment. For educators and organizations already using Moodle, it could be leveraged immediately to capture data to add to existing learning analytics.

The browser extension could be used with student involvement in research efforts by giving them first-person control over data collection, analysis, and potential sharing. It can also be administered through Google Classroom and Workplace centralized controls through Chromebooks replicating a lab-type setting for data collection.

Because writing data captured from "Cursive" has the potential to be drawn from an authentic, un-timed environment, and because it enables both students and instructors to track writing process data over time, "Cursive" offers users a unique way to triangulate qualitative writing process data. In the last half of the 20th century, researchers frequently relied on ethnographic methods, including think-aloud protocols, to better understand writers' internal processes (Emig, 1971; Flower and Hayes, 1981; Rose, 1980). While the empirical data researchers collected via such methods did advance the field's understanding of the writing process, ethnographic data collection methods rely heavily on respondents' ability to convey their experiences reliably.

More recently, efforts have been made to develop writing process models that combine qualitative and quantitative data, particularly that collected via keystroke analysis (van den Bergh et al., 2016; Wengelin et al., 2019). However, many studies continue to be conducted in timed, inauthentic environments that do not lend themselves easily to longitudinal data collection.

Additionally, this approach is designed to be completely transparent, providing writers with access to their data and the analytics and automated scripts and algorithms that are applied to it. Researchers and students alike have the opportunity to review their own beliefs about the writing experience vis a vis the quantitative data collected about that same experience. We believe this approach benefits institutions, instructors, and students by democratizing collection and creating a more transparent writing data collection opportunity. For example, using a tool like “Cursive,” teachers could elicit ethnographic data from students about their writing processes, which students could then compare with the quantitative data collected. We hypothesize that, over time, students who better understand their writing process from *both* a quantitative and qualitative lens will be able to focus on participating in activities and interventions designed to improve specific skills, such as generating ideas or creating logical connections.

Plugin User Experience

Testing of the plugin has been ongoing, including both teachers and students on a dedicated Moodle site with public access and clear statements related to the capture of key log data during registration, upon first login, within activity descriptions, and with end users leveraging the browser extension. All users have remarked on the ease of use and unobtrusive availability of the plugin, with one student remarking that they “didn't think about it at all. It was no different than how I usually write in class” (Hoffman, G., personal communication, October 10, 2023).

Feature Extraction and Analysis

Below is a list of preliminary features easily extracted through data analysis using tools such as Excel:

- Backspace percent
- Character count
- Characters per minute
- Key count
- Keys per minute
- Total time seconds
- Word count
- Words per minute
- Longest Session

Key log data is simple time-series data and amenable to both manual and automated feature development methods. Dozens, hundreds, and thousands of features can be created from key log data depending on the goals of the Machine Learning models (Acien et al., 2022; The Learning Agency, 2023) which demonstrates the robustness of the raw data format.

Data Sharing and the Anonymized “Q-type” Format

By definition, biometric typing data is personally identifiable information (PII; Information Commissioner’s Office, n.d.). While the value of keylog data rests in the calculation of time and distances between known keys, some keylog research, such as that into writing quality, may be achieved by keeping but obfuscating the specific keys in key events. One recent exploration of this occurred late 2023 with the Kaggle Competition focused on Writing Quality (The Learning Agency, 2023). By converting all alphanumeric keys to "q" a new format ("Q-type") was created and shared with competition participants (keys such as 'shift,' 'delete,' 'backspace,' and 'enter' were retained).

Interestingly, the availability of event data, despite being all converted to "q" still allowed for specific and interesting feature extraction and did reliably facilitate deep research into the behavioral elements of writing quality and the simple list of features highlighted above. This format, or one similar, could be easily leveraged as a subset of the data standard described above to facilitate wide research without relying on personal information (keylog data without specific keys is no longer a personal identifier).

Directions for Further Research

Behavioral biometrics is a growing research area. This innovation in code will support researchers and educators by enabling additional data sources and promoting additional research trajectories. The plugin will continue to be updated and expanded to support researchers and educators in research and practical application of existing machine learning models. Several areas of further development include supporting additional device inputs (beyond keyboards), expanding contextual data such as cursor location, adding additional event listeners to create additional data granularity, and expanding the features extracted from the data, including advanced graphical representation (Lam et al., 2021) and calculations such as Shannon's Entropy, Verbosity (Allen et al., 2016), and the standard deviation of interkeystroke intervals (Conijn, Roeser, & van Zaanen, 2019). All development updates will be posted to the public GitHub repository at https://github.com/cursiveinc/cursive-moodle-tinymce.

There are many directions for further research of biometric data itself. We intend to promote this tool as a viable option for key log studies and encourage collaboration. Below, please find a sample of potential research areas underway in various states of maturity across higher education institutions and industries.

Essay Quality

It's estimated that as much as "76% of the variance in the essay scores" (Allen et al., 2016, p. 26) can be attributed to data captured via key logs. Based on these insights and the data features from which they're derived, the potential exists to encourage or guide students to improve their writing in real-time. The editing process can be confusing, esoteric, and abstract to students. Data-driven editing suggestions could assist students by putting effort into the context of specific actions and tasks.

Automated writing evaluation tools (for both scoring and feedback) exist. However, most focus on the finished product rather than the writing process (Baffour et al., 2023). Key log data capture and analysis can ensure that student effort, rather than the finished product, is taken into account in feedback processes, attuning it to a student's needs and goals (Allen et al., 2016).

In fact, in the last 24-months, Kaggle (a subsidiary of Google) facilitated an open competition focused on a large dataset of key log data collected via Amazon's Mechanical Turk, which includes 1000s of SAT-like Essay submissions (The Learning Agency, 2023). The competition, sponsored by the Gates Foundation, Vanderbilt University, and The Learning Agency, hypothesized that essay quality and key log data can be correlated. This underscores the need for infrastructure, such as the plugin introduced here, to support educators, administrators, and technologists in leveraging the insights for feedback and scoring created and published through the competition in natural learning environments.

Identity Verification

One of the most compelling research areas of biometric key log data has already led to significant commercial applications in areas of identity and credit card fraud, account authentication, and spam identification (Acien et al., 2022), the ability to identify and verify users through keystroke data has many uses. User authentication through this means may have significant value in the classroom as well, protecting institutional learning management systems beyond the point of entry (login screens). As Teh, Teoh, and Yue have stated in their survey of various biometric approaches, keyboard biometrics are "unique to each individual and hold huge potential as personal identifier[s]" (2013, p. 2).

Recent innovations in academic integrity have largely focused on AI-classifiers to identify the use of Generative AI in course writing (Ibrahim et al., 2023). The use of these tools is contentious (Fowler, 2023). The proliferation of AI-created content highlights some shortcomings of automated marking and feedback of static text (which may be AI-generated). It is possible that biometric data from keyboards, reflecting the effort of a specific student, may offer a better approach to supporting integrity in classroom writing.

Cognition, Insight, and Feedback

Studies using biometric key log data have highlighted several areas for further research, including changes in individual behavioral biometrics as they are connected to the task being carried out. In one analysis of timed writing, the authors posit:

> More interesting is the presence of some distinct outliers in the timing features. The unusual ways that these students are spending their time may be a symptom of a deeper cognitive difficulty or poor test-taking strategies. Being able to alert teachers to these unusual patterns for follow-up may be a beneficial side-effect of the work on studying timing during writing. (Almond et al., 2012, p. 49)

With biometric data instantly available as students work, insights and interventions can be utilized more quickly and effectively as feedback loops shorten and tighten. Paired with natural language processing-based feedback tools, such as the ThesisWorkshopper (Oenbring, 2022), new frontiers in automated feedback could beneficially support students during the writing process and drive successful behaviors and practices through directed interventions and messaging.

Health and Mood

Finally, there do exist some studies that suggest key log data could be used for health screening or to assist in the identification of mood, cognitive function, or ailments. As a broad research area, health and wellness research related to key log data could be as simple as providing teachers a better sense of student emotional states (boredom, frustration, excitement, engagement, stress, happiness) to improve classroom dynamics and measure student interest (Allen et al., 2014; Fairhurst, Li, & Erbilek, 2014)

At least one study leveraged key log data to identify Multiple Sclerosis against a control population (Lam et al., 2021), suggesting that key log data contains information beyond the above use cases described for further research. The health and wellness of students is a potential area for further study to create additional low-cost screening tools that may identify student health concerns earlier. Biometric data captured in educational settings, paired responsibly and confidentially with student demographic and health data, could be utilized to train new algorithms for screening common diagnoses.

Conclusion

In 1996, a working group of the World Wide Web Consortium (W3C) began building consensus on a new data standard that would allow for easy sharing and communication between websites. This new markup language took years to develop, but it laid the foundation for some of the most important innovations we've come to value in internet communications: XML (World Wide Web Consortium (W3C), 2008). This "eXtensible Markup Language" became a common standard across the web ushering in new functionality and benefits for web-users.

While a JSON format for keylogging may not change the internet as we know it, a common standard could spur innovation by many players to ensure that educational institutions, faculty, students, and writers generally can elect to collect key log data and interface with interchangeable services providing access to a growing number of machine learning models designed to analyze and provide insight about captured data from the writing process which is uniquely created by human beings and our direct interactions with a computer interface; a direct and immediate contradiction to the instantaneously available text from generative AI.

A future where the data format is universally accepted would ensure that all users have access to analysis or tools that could further their understanding on many fronts: language localization, diagnostics for writing practice improvements, automated evaluation, plagiarism detection, authorship and originality verification, blockchain-enabled proof of effort, health screening, cognitive screening, guided practice, intelligent tutoring systems, standardized tests, identity verification,... Shared data may open new avenues for research and exploration into cognition and understanding. Expansions to the standard could also accommodate new methods of human computer interface (touch screen, gesture, or voice).

As Adam Grant said, "writing isn't what you do after you have an idea. It's how you develop an inkling into an insight. Turning thoughts into words sharpens reasoning. What's fuzzy in your head is clear on the page…. Writing is a tool for thinking." (Grant, 2022). For what it's worth, biometric typing data is a tool for the analysis of thinking.

References

1Edtech. (n.d.). *Learning tools interoperability*. 1Edtech. Retrieved July 29th, 2024, from https://www.1edtech.org/standards/lti

Acien, A., Morales, A., Monaco, J. V., Vera-Rodriguez, R., & Fierrez, J. (2022). TypeNet: Deep learning keystroke biometrics. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, *4*(1), 57-70, https://doi.org10.1109/TBIOM.2021.3112540

Allen, L.K., Jacovina, M. E., Dascalu, M., Roscoe, R. D., Kent, K., Likens, A. D., & McNamara, D. S. (2016). {ENTER}ing the time series {SPACE}: Uncovering the writing process through keystroke analyses. *9th International Conference on Educational Data Mining, EDM.* https://asu.elsevierpure.com/en/publications/entering-the-time-series-space-uncovering-the-writing-process-thr

Almond, R., Deane, P., Quinlan, T., Wagner, M., & Sydorenko, T. (2012). A preliminary analysis of keystroke log data from a timed writing task. *ETS Research Report Series*, *2*, i-61. http://dx.doi.org/10.1002/j.2333-8504.2012.tb02305.x

Amazon Web Services. (n.d.). What is JSON? *Amazon Web Services.* Retrieved July 29th, 2024, from https://aws.amazon.com/documentdb/what-is-json/

Baffour, P., Crossley, S., Tian, Y., Franklin, A., Rambis, N., Benner, M., & Boser, U. (2023). The Feedback Prize: A case study in assisted writing feedback tools. *The Learning Agency Lab with Vanderbilt University and Georgia State University.* https://the-learning-agency-lab.com/wp-content/uploads/2023/08/TLA-Lab_Whitepaper_24-Aug-kf.pdf

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing?: A study of research ethics and academic publishing. *Research Methods in Applied Linguistics*, *2*(3), 100068. https://doi.org/10.1016/j.rmal.2023.100068

Chenoweth, N., & Hayes, J. (2003). The inner voice in writing. *Written Communication*. *20*, 99-118. https://doi.org/10.1177/0741088303253572

Conijn, R., Cook, C., van Zaanen, M., & Van Waes, L. (2022). Early prediction of writing quality using keystroke logging. *International Journal of Artificial Intelligence Education, 32*, 835–866. https://doi.org/10.1007/s40593-021-00268-w

Conijn, R., Roeser, J., & van Zaanen, M. (2019). Understanding the keystroke log: the effect of writing task on keystroke features. *Reading and Writing, 32*, 2353–237. https://doi.org/10.1007/s11145-019-09953-8

Emig, J. (1971). *The composing processes of twelfth graders*. National Council of Teachers of English.

Fairhurst, M., Li, C., & Erbilek, M. (2014). Exploiting biometric measurements for prediction of emotional state: A preliminary study for healthcare applications using keystroke analysis, *2014 IEEE*

Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, Rome, Italy, 2014, 74-79, https://doi.org/10.1109/BIOMS.2014.6951539

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365-387. https://doi.org/10.2307/356600

Fowler, G. (2023, April 1). We tested a new ChatGPT-detector for teachers. It flagged an innocent student. *The Washington Post.* https://www.washingtonpost.com/technology/2023/04/01/chatgpt-cheating-detection-turnitin/

Grant, A. [@adammgrant] (2022, July 24). Writing isn't what you do after you have an idea. It's how you develop an inkling into an insight. Turning thoughts into words sharpens reasoning. What's fuzzy in your head is clear on the page. "I'm not a writer" shouldn't stop you from writing. Writing is a tool for thinking [Image] [Xeet]. X.com. https://x.com/AdamMGrant/status/1551208238581948416

Hayes, J., & Chenoweth, N. (2007). Working memory in an editing task. *Written Communication*. *24*, 283-294. https://doi.org/10.1177/0741088307304826

Ibrahim, H., Liu, F., Asim, R., Battu, B., Benabderrahmane, S., Alhafni, B., Adnan, W., Alhanai, T., AlShebli, B., Baghdadi, R., Bélanger, J. J., Beretta, E., Celik, K., Chaqfeh, M., Daqaq, M. F., El Bernoussi, Z., Fougnie, D., Garcia de Soto, B., Gandolfi, A., … & Zaki, Y. (2023). Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Scientific Reports, 13*, 12187. https://doi.org/10.1038/s41598-023-38964-3

Information Commissioner's Office. (n.d.). *How do we process biometric data lawfully?* ICO. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/biometric-data-guidance-biometric-recognition/how-do-we-process-biometric-data-lawfully/

Lam, K., Meijer, K., Loonstra, F., Coerver, E. M. E., Twose, J., Redeman, E., Moraal, B., Barkhof, F. De Groot, V., Uitdehaag, B. M. J., & Killestein, J. (2021). Real-world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis. *Multiple Sclerosis Journal, 27*(9), 1421-1431. https://doi.org/10.1177/1352458520968797

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication, 30*(3), 358-392. https://doi.org/10.1177/0741088313491692

Mintz, S. (2021). *Writing is thinking*. Inside Higher Education. https://www.insidehighered.com/blogs/higher-ed-gamma/writing-thinking

Moodle. (n.d.). Statistics. *Moodle*. Retrieved July 29th, 2024, from https://stats.moodle.org/?lang=fr

MoodleDocs. (2023, October 6). TinyMCE editor. *MoodleDocs*. https://docs.moodle.org/403/en/TinyMCE_editor

Mozur, P., Mac, R., & Che, C. (2022, August 19). TikTok browser can track users' keystrokes, according to new research. *New York Times*. https://www.nytimes.com/2022/08/19/technology/tiktok-browser-tracking.html

Oenbring, R. (2022). ThesisWorkshopper: An automated thesis statement evaluator. *Journal of Writing Analytics, 6*., https://doi.org/10.37514/JWA-J.2022.6.1.02

Pisani, P. H., & Lorena, A. C. (2013). A systematic review on keystroke dynamics*. Journal of the Brazilian Computer Society, 19*, 573–587. https://doi.org/10.1007/s13173-013-0117-7

Rose, M. (1980). Rigid rules, inflexible plans, and the stifling of language: A cognitivist analysis of writer's block. *College Composition and Communication*, *31*(4), 389-401. https://doi.org/10.2307/356589

Roy, S., Pradhan, J., Kumar, A., Adhikary, D. R. D., Roy, U., Sinha, D., & Pal, R. K. (2022). A systematic literature review on latest keystroke dynamics based models. *IEEE Access*, *10*, 92192-92236. https://doi.org/10.1109/ACCESS.2022.3197756

Selfe, C. L., Wahlstrom, B. J. (1988). Computers and writing: Casting a broader net with theory and research. *Computers and the Humanities, 22*, 57–66. https://doi.org/10.1007/BF00056349

Stross, R. (2010, September 4). A strong password isn't the strongest security. *New York Times*. https://www.nytimes.com/2010/09/05/business/05digi.html

Talebinamvar, M., & Zarrabi, F. (2022). Clustering students' writing behaviors using keystroke logging: A learning analytic approach in EFL writing. *Lang Test Asia, 12*, 6. https://doi.org/10.1186/s40468-021-00150-5

Teh, P. S., Teoh, A. B. J., & Yue, S. (2013). A survey of keystroke dynamics biometrics. *The Scientific World Journal*, 2013, 24. https://doi.org/10.1155/2013/408280

The Learning Agency. (2023). Linking writing processes to writing quality. *Kaggle*. https://www.kaggle.com/competitions/linking-writing-processes-to-writing-quality

TinyMCE. (n.d.). Events - TinyMCE 6 Documentation. *TinyMCE*. Retrieved July 29th, 2024, from https://www.tiny.cloud/docs/tinymce/6/events/

van den Bergh, H., Rijlaarsdam, G., & van Steendam, E. (2016). Writing process theory: A functional dynamic approach. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 57–71). The Guilford Press.

Wengelin, Å., Frid, J., Johansson, R., & Johansson, V. (2019). Combining keystroke logging with other methods: Towards an experimental environment for writing process research. In *Observing writing* (pp. 30-49). Brill.

World Wide Web Consortium (W3C). (2008, November 26). *Origin and goals*. Extensible Markup Language (XML) 1.0 (Fifth Edition). https://www.w3.org/TR/REC-xml/#sec-origin-goals