PoetryWordle

(Technical Paper)

Optimizing the Usage of Video Games in Youth Education

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree

Bachelor of Science, School of Engineering

Noah Holloway

Fall, 2021

Technical Project Team Members

Noah Holloway

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature ____ Date __________

Noah Holloway

Approved ____ Date __________

Nathan Brunelle, Department of Computer Science

Approved ____ Date __________

STS Travis Elliott, Department of Engineering and Society

The art of poetry has always been at once fascinating and frustrating to me. As someone who believes he is creative, I am fascinated with the concentric layers of beauty that can densely fill a poetic work, leaving something new to be discovered with every read. As an engineer, I feel oddly at home with the constraints provided by a poetic framework, and have enjoyed seeing how the literary masters have sculpted their work to fit, or break, these rules, as well as trying my hand at writing poems to fit these constraints myself. It's rather trivial to come up with sets of rhyming words, or a three syllable word with the second stressed to fit some alphabet soup of a rhyme scheme, but to do so in a way that still yields an inspiring poem is a much greater feat. At times, however, this freedom of expression can feel too vast, and the map to traverse the realm of poetic understanding can feel outdated. The constraints in writing poetry are somehow both numerous and sparse, there are somehow no wrong answers, yet very few right answers, making poetry both a trivial and impossible optimization problem! Barring one embellished for advertising purposes, there is no class titled, "How to Write Exactly Like Shakespeare." In fact, why are Shakespeare's centuries-old works the universal choice to study in English classes, not amateur works in vast online repositories that are arguably far more pertinent? When the standard of comparison is this arbitrary, it can be difficult to teach and measure mastery to novice poets. Distinguishing expert and novice poems is in fact a lot like a contemporary topic in AI, that of distinguishing human-generated text from that produced by a computer. Fagni et al., as part of their TweepFake project, compiled a list of real and fake Twitter posts, and ran various detection algorithms on the dataset, to determine how accurately the posts could be categorized (2021). On the whole, their detection was more accurate than not, but text snippets generated by more novel, state-of-the-art models tended to be more challenging to classify. If separating man from machine is so challenging, how do we separate man from man?

The technical project aims to directly address this issue, by using a machine learning approach trained on works from famed poets over history. The insights gathered here could serve as a rubric, capable of measuring the effectiveness of new works by comparing them to works of the past. In turn, the STS research project closely inspects a methodology taken for granted in the technical project, namely the strategy of using video games to enhance the learning experience of students. It is important to consider these two questions in tandem, rather than separately. An impressive technical demonstration is of little value if applying it in the classroom would cause preventable social or technical issues.

DESIGN OF AUTOMATED TOOL TO ALLOW FOR INTERACTIVE POETRY INSTRUCTION

Standardizing grading is difficult. There are several schools of thought when it comes to scoring even the most discrete, clear-cut assignments, let alone poems written by students. When it comes to the quantitative result, some propose grading on a curve, relative to one's peers, where others propose grading every submission as if it were in a vacuum. Ladas (1974) defines this latter style as "competency grading." On a competency grading scale, there should be "an explicit statement of goals formulated" and "agreement should be reached among instructors" about the goals that are being measured, and the goal should be to assign scores based solely on the contents of the submission, not that of other students' works (Ladas, 1974). How, then, do we explicitly "formulate the goals" needed to use such a grading style? We could attempt to mimic the grading styles of the best teachers, but delineating why they are "better" can be equally as difficult. Are teachers who award higher scores on average "better?" Maybe they are being too generous, so what about teachers with lower averages? This may be more realistic, but is trivialized by the extreme of awarding zero percent to every student on all assignments. Lower average scores may also demotivate students, discouraging them from putting in effort for fear that their work may not be appreciated.

Rather than trying to piece together partial solutions from studies of human graders, we can turn to a more self-sufficient solution, in the form of automatic graders, or "autograders." Autograders have been routinely implemented in programming classes to reduce the amount of work left to course staff. For instance, a submitted Java program may be automatically run with a set of inputs, then graded on how many inputs produce the right output. The simplest form of an autograder, of course, is an automated test-taking software that restricts the student's inputs to multiple choice or short-form answers, which can be compared for precise equality. A correct selection out of four possibilities earns full points, an incorrect one earns none. This requires no inference, and can be done with certainty. There are two main benefits to using an automatic grader, especially in a formative (rather than summative) context. First, it reduces the workload on the grader, and (of course) improves consistency between submissions. Secondly, it allows for immediate feedback, supporting student progress (Lin, 2019).

This project, in a similar sense, will serve as an autograder for poems. A naive approach might be to simply count how many words in poem A are in the same position in poem B, but this is a rather unenlightening metric, and is susceptible to wild changes in estimated inaccuracy when a word is appended to the front of the poem, for instance. Rather than giving a raw score out of one hundred percent, this tool will compare features of two poems against each other, and award similarity scores based on these features. For example, it might say that while the rhyme scheme of the poems matches exactly, the meter only matches sixty percent of the time. Once these features are extracted and separated they become incredibly versatile, and can be used to construct various educational models, which are discussed in detail later. Simply having this insight about relative similarity of two poems is the vast initial hurdle, and the methodology is critical.

This project will be conducted under the direct supervision and expertise of Nathan Brunelle, an assistant professor in the Department of Computer Science at the University of Virginia. Additionally, while not a direct supervisor, Professor Brad Pasanek of the University of Virginia Department of English will be a tremendous resource in answering questions related to the mechanics of poetry-writing, as well as providing direction to poems and authors of significance, in order to hone the data acquisition for this project. This project is not presently being conducted with any explicit team members, but the Puzzle Poetry group, organized by Professors Brunelle and Pasanek, will be consulted frequently to discuss ideas and progress on the project. This group holds office hours weekly on Fridays, and consists of several other graduate and undergraduate students working on cutting-edge poetry-related projects. The timeline of this project is specified in the Gantt chart in Figure 1 (Holloway, 2021).

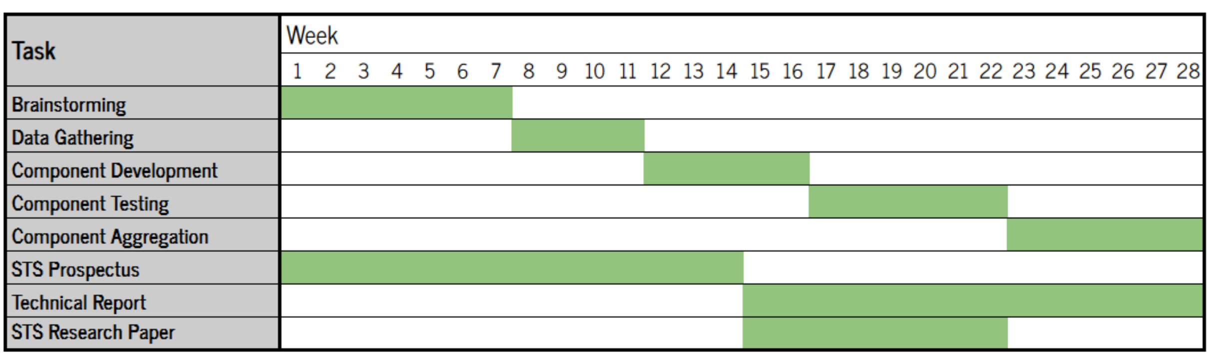

Figure 1: Gantt Chart. This figure shows the planned work to be achieved over the remainder of the 2021-2022 academic year, for the Technical and STS Capstone. Cells in green signify work that is forecasted for that week. (Noah Holloway (2021)).

The design process used to achieve the project objectives will consist of 1) define the problem, 2) specify requirements, 3) acquire data, 4) generate concepts, 5) design, 6) test and validate, 7) apply results. This is represented in Figure 2 (Holloway, 2021).

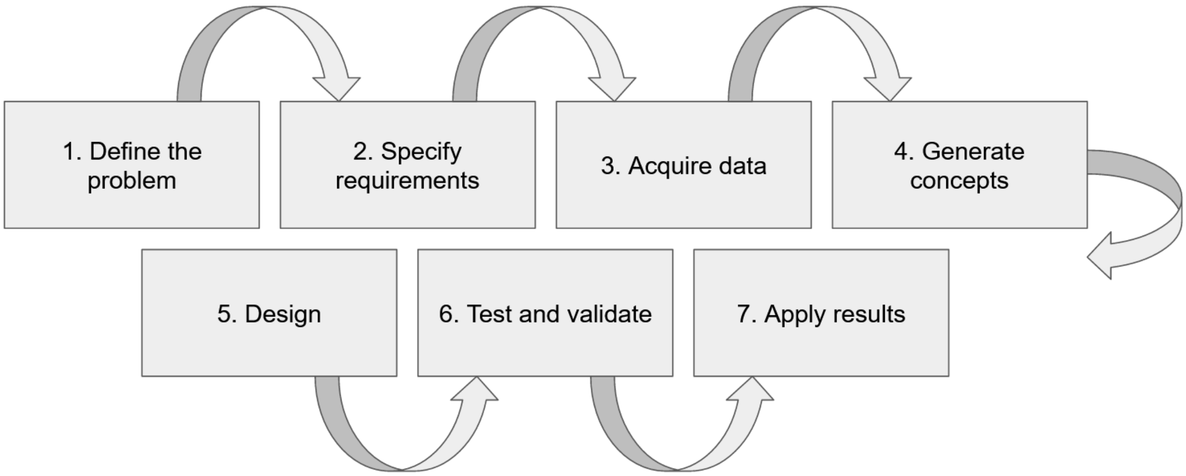

Figure 2: 7-step Forecasted Process. This figure shows the anticipated process for developing the technical application. The data referenced in step 3 will consist of a set of poems for training a machine learning algorithm. The results in step 7 will consist of some interesting presentation of the features discovered. (Noah Holloway (2021)).

The expected output from the conclusion of this product is a user-facing application that takes the features determined through the machine-learning approach and uses them as instructional baselines for students to practice writing poems. The exact abilities of this application remain uncertain and will be determined based on the poem features that are able to be extracted from the dataset, and how expansive the dataset proves to be. English Professor Brad Pasanek, aforementioned member of the "Puzzle Poetry" group, will be invaluable in determining which metrics are relevant to examine. These metrics will likely include rhyme scheme, sentiment, intensity, among others. For each feature, the goal is to execute a form of supervised (labelled) machine learning on a large dataset of similar words and phrases. For example, to map happiness versus sadness across a poem, our algorithm would first peruse thousands of English words, and be told whether those words are happy or sad. After enough training, the goal would be to predict the relative "happiness" of a word or word sequence, and to repeat this prediction across the entirety of a poem. Francisco & Gervás (2012) present their work with sentiment analysis in EMOTAG, and automatically marking up text as one of many emotions of varying intensities, which is exactly the type of categorization sought after here. Liu (2020) also discusses running sentiment analysis on short product reviews, as a simple case. This is a good model to start out with as it is not nearly as complex as a fourteen-line sonnet.

Once complete, the tool will "quiz" the user to recreate a deconstructed poem. The features extracted will be presented in full, and the user will be tasked with combining the several different contexts at their disposal to reassemble the puzzle. Foreman describes a somber personal anecdote of attempting to solve a nine-thousand piece puzzle with his family (2015), but highlights how he at times got entirely lost in the allure of finding a solution. This tool will replicate this same captivation of solving a physical puzzle, by reaffirming the user when they "click" a word into place and offering little guidance when they are incorrect. As de Aguilera & Mendiz (2003) say in the context of video games, "the immediate feedback provided… challenge and stimulate children and adolescents and arouse curiosity, which can be extremely useful in learning." An advanced iteration of the tool could even prompt the user to construct an entirely new poem, modeled after an existing one. Rather than providing a "bank" of words as puzzle pieces, the slate would be wiped entirely clean, letting the user write freely, and the user would occasionally be told whether the derived features of their own work match the hidden background features of another's. Such a goal is lofty but would be a partial improvement on the work done by Gervás et al. (2007) in automatically generating poems. They did not face much success, but perhaps by involving a human participant (or rather, by having a human write with a computer assistant), they can generate works that neither could individually.

OPTIMIZING THE USAGE OF VIDEO GAMES IN YOUTH EDUCATION

There is growing debate over the extent to which technology should play a role in the classroom. One example of this is the rapid increase in the frequency of lessons delivered virtually, over a conference-call platform, mainly in response to social distancing demands brought about by the COVID-19 pandemic. Virtual classrooms, while effective at limiting the potential spread of the novel virus, come with their own host of issues, to include watering down the educational experience for students who have now overworked teachers, to equitability concerns for those families who do not have the resources to join online classes, due to lack of access to a computer or consistent internet connection (Abril, 2021).

Can we use video games in the classroom to deliver instructional content? What about managing student behavior? Can we do both at once? These are relatively novel questions, but they are extremely worth exploring, given that my technical project is quite literally supposed to be a computer game for teaching poetry in the classroom. As our schools become more digital, it may seem only natural that games and gaming becomes an integral part of our school curriculum, and there are voices for the change, as well as steps being taken in that direction. There are, however, some potential side effects of making this tech so prevalent, to include a risk of alienating those individuals who cannot afford consistent access to the technology, and the issue of perpetuating a growing trend of video game addiction.

According to Schaaf and Mohan (2014), students "go hungry in schools" intellectually, and attempt to learn "in a disconnected fashion," when contrasted with the arguably less meaningful lessons they learn at home, playing video games while online with friends (p. 2). They ask rhetorically, "what if teachers were to utilize the same technologies in the classroom?" (p. 2). One possible explanation is that students learn the same amount of a topic in a given amount of time, but since video games are more stimulating, they are more likely to spend more time in game, and thus learning from it. There seems to be quite a bit of untapped potential of which games with a central focus on some otherwise archaic educational topic could make use. On the other hand, playing video games has been shown to activate "similar brain regions as drugs of abuse, including the mesolimbic reward system and amygdala" (Mathews et al., 2019). Video games are addictive, and must be handled with care. It becomes essential to investigate the degree of risk inherent to a naive universal implementation of games in schools. What are the problems that may arise from ubiquitously implementing instructional video games? Are there more measured approaches to take, in order to make use of the benefits provided by this novel teaching style, while taking into account many of the potential risks involved?

The answer to this final question should be an affirmative one, and is the stated goal of this research: to empirically outline a set of guidelines for academic "gaming," of various sorts, ranging from the commonplace live, point based "pop quiz" through popular internet apps, to the extreme scenario of orienting an entire lecture or even class solely around mastering some mechanically and intellectually stimulating game, despite it failing to meet any explicit educational standards. Answering this question will necessitate an in-depth literature review, with technical accomplishments and data tending to buttress a pro-gaming perspective, and humanistic and societal analyses arguing against. Where possible, I will embed my own experience both as a student and someone who has played video games in the past, not as the foundation for my analysis but as an illustrative anecdote where appropriate. The chief framework for analyzing this discussion will be Actor-Network Theory (ANT), one that emphasizes viewing relationships among actants from the ground up, and making very deliberate choices as to which actors are relevant to the analysis. Without using ANT, it becomes infeasible to "inspect the precise ingredients that are entering into the composition of the social domain (Latour, 2005)," and it is exactly these precise ingredients that will form the basis of the analysis. There will be a scrupulous focus on the primary classroom actors, namely students and their teachers, and secondary actors may be incorporated from the students' own immediate networks, including social relationships with peers and familial relationships with parents.

One naturally useful distillation of ANT to analyze this debate is the Technology and Social Relationships Model. In this model, we analyze the effects that the end user has on other actors as a result of acquiring their new technology. In this specific instance, the end user is the student learning with the aid of a video game tool. The other individuals being affected may include their teachers, parents, and other students in the class. This specific instance is shown in Figure 3 (Holloway, 2021). In this way, we can examine the totality of positive and negative influences on other actors in the student's educational network, in order to more quantitatively determine the viability of implementing this tech.

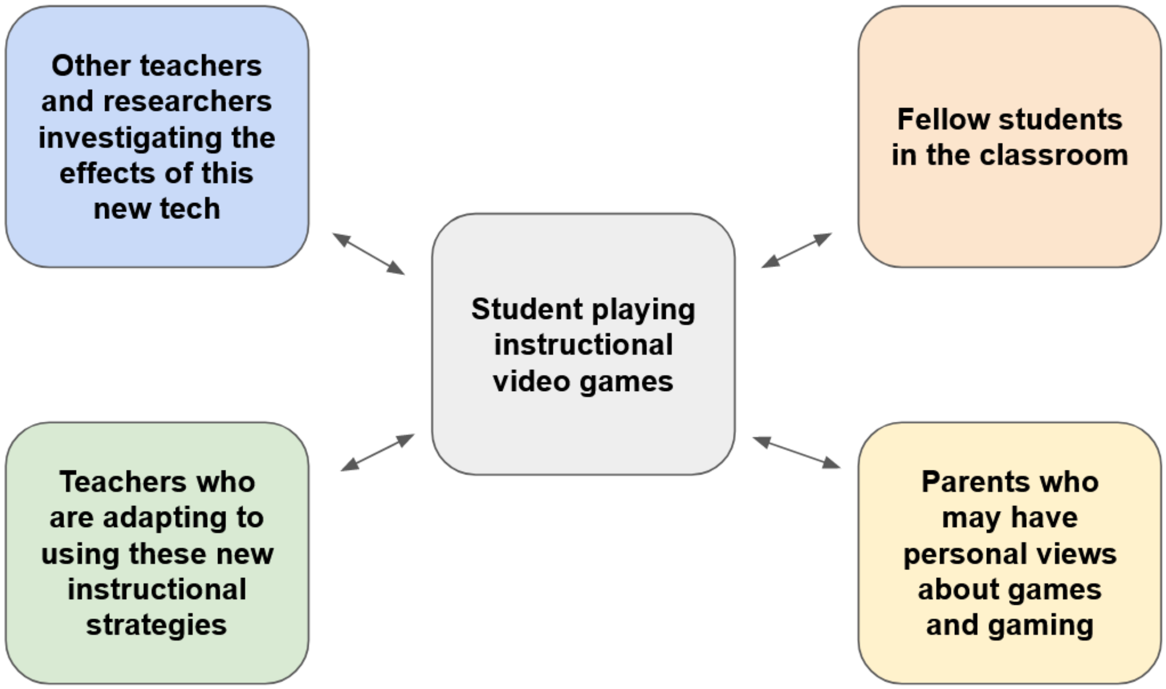

Figure 3: TSNM and Video Games. This figure shows the Technology and Social Networks Model applied to the usage of educational video games. The student, using the game, appears in the center, and some possible individuals in his/her local network appear on the outskirts. (Noah Holloway (2021)).

This research will be presented in the form of a scholarly article, with the goal of delineating when it is appropriate to incorporate these new, exciting instructional strategies into the classroom, and when more traditional strategies may be appropriate. The goal is not to conclude definitively whether one is better than the other, and it is expected for there to be some level of nuance in the answer. The intent is to provide some level of guidance to both inform my attempts to publicize and apply the aforementioned technical project, and to inform others embarking on similar projects in the future. If some circumstances are found to be unwelcoming to this sort of change, the goal is to recommend further avenues of research in order to ensure that students of the future are learning as efficiently and fairly as possible.

As the scope of the examined network extends, broad issues will arise, and will need to be addressed in turn. For instance, Lachney et al. (2018) discuss standard ethical issues that come with using technology in the classroom, issues which extend far beyond the room's four walls. One part of their five-pronged "Educratic Oath" is that educators should do "nothing to impair learning, performance, and instruction." They then provide the unexpected example of a teacher using computers, incidentally, from complains that use child labor to manufacture parts. By using the fruits of this labor, labor which is taking children away from learning and instruction, it can be argued that the teacher is violating this tenet of the Oath. Obviously, it will be impossible to predict every possible way these fundamentals could propagate through the examined networks, but an acknowledgement of the potential far-reaching side effects is necessary.

This research is not only extremely significant, but urgent. Technology is advancing at a blistering pace, and technological dependence is growing just as rapidly. Herein lies a fitting callback to the optimization problem referenced in the Introduction, as this presents another crucial optimization. We must strive to enrich the learning experiences of today's young students, making it easy and accessible for them to learn *with* technology, while avoiding a scenario such that they cannot learn *without* technology.

References

Abril, D. (2021, October 4). Back in the classroom, teachers are finding pandemic tech has changed their jobs forever. *The Washington Post.* https://wapo.st/3pxHDG5

de Aguilera, M., & Mendiz, A. (2003). Video games and education. *Computers in Entertainment*, *1*(1), 1–10. https://doi.org/10.1145/950566.950583

Fagni, T., Falchi, F., Gambini, M., Martella, A., & Tesconi, M. (2021, May). *TweepFake: About detecting deepfake tweets*. PLOS ONE. Retrieved September 23, 2021, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0251415.

Foreman, R.L. (2015). Jigsaw Puzzle. *River Teeth: A Journal of Nonfiction Narrative 17*(1), 25-27. doi:10.1353/rvt.2015.0027.

Francisco, V., & Gervás, P. (2012). EMOTAG: An approach to automated markup of emotions in texts. *Computational Intelligence*, *29*(4), 680–721. https://doi.org/10.1111/j.1467-8640.2012.00438.x

Gervás, P., Hervás, R., & Robinson , J. (2007). *Difficulties and Challenges in Automatic Poem Generation: Five Years of Research at UCM*. https://www.researchgate.net. Retrieved September 23, 2021, from https://www.researchgate.net/profile/Pablo-Gervas/publication/228472586_Difficulties_and_Challenges_in_Automatic_Poem_Generation_Five_Years_of_Research_at_UCM/links/02bfe50cae4d026d9f000000/Difficulties-and-Challenges-in-Automatic-Poem-Generation-Five-Years-of-Research-at-UCM.pdf.

Holloway, N. (2021). Gantt Chart.

Holloway, N. (2021). 7-step Forecasted Process.

Holloway, N. (2021). TSNM and Video Games.

Lachney, M., Boltz, L. O., Dillman, B., Robertson, C., & Yadav, A. (2018). Local Classrooms, Global Technologies: Toward the Integration of Sociotechnical Macroethical Issues Into Teacher Education. *Bulletin of Science, Technology & Society*, *38*(1–2), 13–22. https://doi.org/10.1177/0270467620902972

Ladas, H. (1974). Grades: Standardizing the Unstandardized Standard. *The Phi Delta Kappan*, *56*(3), 185–187. http://www.jstor.org/stable/20297851

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Lin, K. (2019, May 28). A Berkeley View of Teaching CS at Scale. arxiv.org. Retrieved November 1, 2021, from https://arxiv.org/pdf/2005.07081.pdf.

Liu, B. (2020). The Problem of Sentiment Analysis. In *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions* (Studies in Natural Language Processing, pp. 18-54). Cambridge: Cambridge University Press. doi:10.1017/9781108639286.003

Mathews, C. L., Morrell, H. E. R., & Molle, J. E. (2019). Video game addiction, ADHD symptomatology, and video game reinforcement. *American Journal of Drug & Alcohol Abuse*, 45(1), 67–76.

Schaaf, R. L., & Mohan, N. (2014). *Making school a game worth playing : Digital games in the classroom*. SAGE Publications.