Original Research Article

Mathematical Models in Quantifying Personal Productivity Among High School Students

Guhan S. Gargya

Obra D. Tompkins High School, 4400 Falcon Landing Blvd, Katy, TX 77494, United States

ABSTRACT

While there is literature on defining personal productivity and connected factors, no studies have covered the personal productivity of high school students, nor found a method of quantifying it. The purpose of this study was to fill this gap and study productivity among high school students to inform the process of creating a mathematical model to quantify personal productivity. Data was collected from 10 participants using a quantitative method, with participants answering Math and English questions modeled after high school Algebra I, Geometry, and Digital SAT content over three separate sessions. After each assessment session, participants were asked to self-score their Perceived Personal Productivity (PPP) on a scale from 1 to 10. The data suggested that the main components of high school students' productivity were amount, accuracy, and efficiency, and the relationship between PPP and actual productivity showed weak correspondence. Drawing from these preliminary findings, a model was created which relates the relevant factors of productivity.

Keywords: personal productivity, efficiency, amount, perception, mathematical model, validity

INTRODUCTION

In today's world, managing personal productivity–a person's management of time, energy, and resources to efficiently achieve their goals and complete their tasks–has become a much bigger challenge. With the rise of technology there are so many more interruptions and external factors that take effect on an individual's productivity, and thus, personal productivity has evolved into a more complex concept.

Because of this change due to modern circumstances, there are many more complex problems associated with personal productivity (1, 2). This is a problem because not only do these modern challenges directly impact one's personal productivity, but it also impacts the actual actions that one takes. High school students have dealt with these effects, as interruptions from phone notifications, watching long and short-form content, and other external factors diminish their focus and efficiency while working or switching tasks altogether.

In order to grasp the idea of personal productivity, it is crucial to find its definition. In an industrial or business context, productivity refers to a relationship between measured outputs and inputs which is typically represented as productivity equals the total outputs divided by the total inputs (3).

While there are other adaptations to the formula, as summarized in Table 1, they all mainly follow this standard relationship.

While this definition of productivity doesn't match the definition of personal productivity exactly, the general idea of evaluating the relationship between output and input can be applied when quantifying personal

Corresponding author: Guhan S. Gargya, E-mail: guhan.sgargya@gmail.com.

Accepted July 17, 2026
https://doi.org/10.70251/HYJR2348.44364379

Table 1. *List of different adaptations of productivity formulas in industrial and business contexts.*

Productivity formulas
Productivity = outputs + inputs
Productivity = actual output + expected resources used
Productivity = (outputs + inputs) × quality
Productivity = efficiency × utilization × quality
Productivity = (output quantity × output quality) ÷ (input quantity × input quality)

Note. "Productivity" is defined here as the measure of economic performance that compares the amount of outputs with the amount of inputs used. "Outputs" refer to the amount of an item or service produced; "inputs" refer to resources required to produce outputs. "Efficiency" in this context refers to how effectively results are produced in relation to time, "utilization" refers to the extent that resources are used, and "quality" refers to the standard of the produced item or service. Data adapted from Overview and classification of approaches to productivity measurement (3).

productivity as well. Kim and colleagues designed a two-week case study experiment that used diary entries to track the productivity of workers which helped them to break personal productivity into four main sectors: amount, a numeric quantity that represents how many of a certain task or metric was completed; efficiency, the amount of work done in relation to the time spent; feeling, a person's physical state, emotional state, and attitude toward their work; and alignment, the long-term benefit of the completed work (2).

While the four main aspects define productivity, they don't affect productivity itself. Borst and colleagues used a computational model to research the effects of interruptions on cognitive processes and identified a key factor known as the "problem state bottleneck," which states that interruptions which challenge or push the brain have a severe adverse effect on the ability to resume tasks, affecting productivity (1). Similarly, Yi and colleagues analyzed over three million records from a manufacturing assembly line and found that productivity increases in the span of time preceding a planned break and decreases right after a break from work, with these results becoming more extreme with the increase of workload (4).

It is also necessary to understand how to effectively measure productivity, because a wrong approach could result in inaccurate findings. Gennara, in their discussion of how the perception of productivity should be factored when measuring personal productivity, validated the Subjective Productivity Scale (SPS), which takes into account perceived effectiveness, perceived efficiency, and perceived focus in order to measure productivity by a perception-based scale (5). These qualitative measures of productivity, along with more common quantitative measures, are implemented in productivity-tracking apps. For example, Kim and colleagues developed an application known as OmniTrack, a flexible tracking tool that people can enter any aspect of their personal data into without any restrictions, and includes important productivity features like a Pomodoro timer, which allows for productivity measurements to account for a higher variety of factors, as well as more qualitative factors (6, 7). However, not all scholars believe in the measurement of productivity. Ko argues that simple measures of productivity fall victim to Goodhart's law, which states that when a measure becomes a target to reach, it stops being a good measure, and thus, these one-dimensional, basic measures should not be implemented (8).

While there is a large breadth of literature on the many robust ways to measure personal productivity, and on the changes people can make to improve their personal productivity, there is a gap in the existing material about integrating these two concepts of quantification and optimization, especially in regards to students. Considering that there is an established negative association between an interruption's cognitive complexity, and the capacity to resume a task, there should not only be an implemented method to measure the change in productivity, but also a way of understanding how to improve productivity in this aspect (1). This idea holds true for all aspects of productivity, as it is not enough to take a measure, but also to understand the approach to improving productivity in that specific sector, in order to help people, particularly students, with productivity issues.

With the identified gap in the field of personal productivity and its measurement, this research seeks to answer the question: how can a mathematical model be constructed to quantify personal productivity among high school students? To answer this question, a quantitative experimental approach was designed to assess the key components of personal productivity among high school students.

METHODS AND MATERIALS

In order to answer the overarching question, a mainly quantitative approach that thoroughly assessed the productivity of students under varying scenarios–in order to assess three of the four main aspects of productivity (amount, efficiency, and feeling)–was used (2).

Research Design

The study employed a mainly quantitative approach that used quantitative data collected from an assessment website, as well as perception-based data from the participants. The first part of the study revolved around collecting quantitative data through a series of controlled Math and Reading assessments over the span of two hours, with three sessions that had varying conditions: one being normal, one having a planned break, and one having random sudden interruptions. The second sector of the study was based on a survey administered after each work session that assessed each participant's perception of their productivity during that respective session. The survey was an adaptation of the SPS, which was validated by Gennara, with this survey measuring Perceived Personal Productivity (PPP) on a scale from 1 to 10, instead of 1 to 7 like the SPS (5).

Research Methodology

Among the existing literature on productivity, studies commonly employ quantitative, or mixed-methods of study in order to determine important aspects of productivity such as efficiency and

amount (9, 10). The approach that was used can be justified by Kim and colleagues, who define four main aspects of productivity, three of which this study analyzed: amount, efficiency, and feeling (2). Because the study aimed to assess these values, it was important to use both quantitative methods–for amount and efficiency–and qualitative methods–for perception–that are translated into quantitative methods. This combination of methods of assessment is also supported by multiple authors, who argue for an integrated approach of subjective and objective measures when measuring productivity (5). In addition, Gennara's study confirms the validity of the SPS model for measuring perceived productivity; by editing the scale to have a broader range, participants can have more diverse answers, leading to more potential for data analysis. This combination reveals more information in comparison to singular objective methods, and it is more robust and valid, reducing the risk of oversimplification, which is common in productivity research (3, 5).

Operationalization of Variables

To ensure replicability, the key aspects of productivity from Kim *et al*. were operationalized. Amount was operationalized as the participant's Total Questions Attempted (Q) , which quantifies the volume of work completed. Total Accuracy (TA) was operationalized separately as the percentage of questions answered correctly out of the total number of questions attempted, representing the quality dimension of productivity. This distinction aligns with Kim and colleagues' framework, where amount refers to the quantity of completed tasks, while quality is a separate consideration (2). Efficiency was operationalized as the Total Questions per Time, represented by the variables Quantity (Q) divided by Total Time (TT) (Q/TT), which was calculated by dividing the number of attempted questions by the average time per question. Thus, a higher value is indicative of a higher speed, which is a crucial component of efficiency. Feeling was operationalized as the participant's PPP, which was recorded from participants' self-reported score on a 10-point Likert scale which follows the SPS (5).

Subjects of Study

A convenience sampling approach was employed to recruit participants from Obra D. Tompkins High School. All participants were randomly chosen across ethnic, gender, and age groups from a set of randomly selected school classes. Informed consent was obtained from participants over the age of 18 and from a parent/guardian for participants under 18. The study was approved by the school's Institutional Review Board. The sample size of 10 participants was selected primarily due to the feasibility and logistical constraints associated with the study's experimental design, especially the extensive testing time (which totaled over an hour per participant). While this sample size is small and limits the statistical power of the findings, it is a justifiable starting point for this exploratory research, as the primary aim of this study was to gather preliminary data and establish the foundational relationships necessary for model construction, and the sample size allowed for a depth of analysis that was sufficient to draw the necessary conclusions (11).

Instruments

Participants took an assessment on a website that contained a combined math and reading assessment, which was controlled to ensure that students were assessed fairly. The website reported specific statistics regarding each participant's performance: number of problems answered, number of problems solved, starting and ending time, and time per problem. This also allowed for graphical data analysis, such as the time per problem graphed in relation with the progression of time during the session, and the amount of problems attempted and solved in relation with the progression of time. Following each session, students completed the aforementioned modified version of the SPS on a Google Forms survey.

Procedure

The study followed the process of collecting quantitative data from an assessment website, and a post-assessment survey. Following this, the data were analyzed. For each of the sessions, participants began with taking the online assessment while undergoing the varying explanatory variable. The assessment consisted of controlled math questions (modeled after high school Algebra 1 & Geometry) and English questions (modeled after the Digital SAT). After the completion of each assessment session, participants were prompted to complete a survey that assessed each participant's perception of their productivity.

Data Analysis

All statistical analyses were performed using Python (version 3.12) with the pandas, numpy, and scipy.stats libraries. The core productivity score was computed using the empirical model derived from the experimental data (Figure 7). For each of the three sessions (no break, planned 1-min break, sudden interruption), the mean productivity score was computed across all participants. The percentage changes between sessions were calculated to test whether the model captured the expected monotonic decrease in productivity as interruptions were introduced. To assess the relationship between objective model scores and subjective self-ratings, both the Pearson product-moment correlation coefficient and the Spearman rank correlation coefficient were calculated between the model scores and the PPP scores for each session individually, as well as for all sessions pooled together. The significance was determined using the corresponding p-values, with a threshold of $p < 0.05$.

Validity Analysis

Five specific validation tests were performed on the collected experimental data to evaluate the stability and consistency of the model. Firstly, through the Mean Score Comparison test, the mean model scores for Session 1 (no breaks), Session 2 (1-min break), and Session 3 (sudden interruptions) were compared. A decrease ($S1 \geq S2 \geq S3$) was expected, as prior research has demonstrated that interruptions and breaks can disrupt workflow and reduce productivity. Borst and colleagues showed that cognitively demanding interruptions impair task resumption, while Yi and colleagues found that productivity declines immediately following planned breaks (1, 4). Thus, the mean score comparison served as an initial check of whether the model's outputs aligned with these established theoretical patterns. The second test examined whether the new model scores diverged from the participants' subjective perceptions of their productivity. To do this, the Pearson and Spearman correlations between the model output and the participants' PPP scores, both per session and pooled, were computed. Thirdly, Parameter Sensitivity Analysis was implemented to assess the model's robustness which involved systematically varying each of the three parameters (α, β, δ) across values of 0.5, 1.0, 1.5, and 2.0 while holding the others at 1.0. For each parameter value, the mean model score and the Spearman rank correlation coefficient were computed to compare the ranking of participants with the baseline parameter set, where all parameters have a value of 1.0. A strong correlation (close to 1) would indicate that the model's ordering of participants is largely insensitive to the specific parameter choices. Through the Break Penalty Validity test, the effect of break time was isolated by using each participant's average accuracy and average total time across all three sessions, which was used to verify that the break penalty factor (T_{break}/TT; the proportion of the session spent on a break) was applied consistently. The model score was then computed for the same participant under three hypothetical break durations (0, 1, and 2 minutes). Each participant was checked to see if their scores strictly decreased with increasing break time, and the proportion of participants showing this decrease was reported. Finally, the Rank Stability test was important to evaluate whether the model produced consistent rankings of participants regardless of the session condition. This was done by ranking participants by their model score within each session, then calculating the Spearman rank correlation coefficient between the rankings of every pair of sessions (S1 vs S2, S1 vs S3, S2 vs S3), as well as the average correlation across all three pairs. A higher average correlation value indicates that the model produces stable relative assessments of participant productivity. All reported p-values are two-tailed.

RESULTS

Of the 10 participants who agreed to participate in the study and signed the given consent form(s), all 10 completed all three sessions of the experiment without compensation and were tested on January 19th, 2026 over online meetings.

Consolidated Findings

With 10 participants, three sessions per participant, and 30 to 45 questions per session, there were well over 1000 points of data to analyze. The gathered data were split into multiple sections to allow for easier analysis. Mainly, each participant's average time per question, and their accuracy was measured across all three sessions. These accuracy and time values were also blocked into sections of math, English, and totals, to lower the variability and ensure there were no discrepancies that could lead to misinterpreted data. Along with the findings from the assessments, each participant's self-rated PPP score was included to allow for a comparison between subjective and objective measures. While these data points were separated by session, separately, the data from each participant was also unified across the sessions in order to ascertain an overview of each participant's productivity findings. Separate from the findings organized by participants, there was data collected about the questions themselves, analyzing the differences between the trend of time per question, per each session.

PPP Outcomes

Each participant was asked to report their PPP score at the end of each of the three sessions. The sole PPP data was analyzed by averaging each of 10 participants' self-rated PPP score for each session (Table 2).

<u>Session One</u>

Session one of the experiment involved 15 minutes of uninterrupted working time, where participants answered math and English questions. The results, separated by each of the participants, are summarized in Table 3.

The data from session one showed that participants had similar results, except for P6, who had both the lowest accuracy and the longest time per question by a large margin. Also, P3 gave themself a notably low PPP score relative to their performance (Table 3). The consistency of participants' time per question during this session is illustrated in Figure 1, and the relationship between participants' self-reported PPP scores and their calculated formulaic productivity scores is shown in Figure 2.

<u>Session Two</u>

In Session two of the experiment, during the shift from math-based to English-based questions, there was a planned one-minute break for every participant. The results, separated by each of the participants, are summarized in Table 4.

The data from this session revealed very similar outcomes, with participants having performed at the same level in terms of time and accuracy, and again P6 having both the lowest accuracy and the longest time per question–this time by an even larger margin than session one. However, P6 gave themselves a substantially lower PPP score for this session, despite similar results to session one. Similar to the previous session, P3 gave themselves an extremely low PPP score of 1, which was the lowest score of the entire experiment (Table 4). The consistency of participants' time per question during this session is illustrated in Figure 3, and the relationship between participants' self-reported PPP scores and their calculated formulaic productivity scores is shown in Figure 4.

Table 2. *Average Perceived Personal Productivity Score by Session on a 10-Point Scale. Ten high school students participated in the study. Subjective Perceived Personal Productivity (PPP) scores were collected immediately after each session using a modified 10-point version of the Subjective Productivity Scale (SPS). On average, participants rated the uninterrupted first session as the most productive. Mean PPP scores then declined in the second session, which included a one-minute planned break, and again in the third session, which included a sudden interruption.*

Session	Average PPP Score out of 10
Session one: No Interruptions	8.1/10
Session two: Planned Break	7.3/10
Session three: Sudden Interruption	6.5/10

Table 3. *Consolidated Participant-Level Data from Session 1.*** *ET indicates the average response time per English question; EA, overall English accuracy; MT, average response time per mathematics question; MA, overall mathematics accuracy; AT, average response time across all questions; and TA, overall accuracy across all questions. P1–P10 represent Participants 1–10.*

Participant	*ET*	*EA*	*MT*	*MA*	*AT*	*TA*	*PPP*
P1	20.000	100.00%	7.067	100.00%	14.457	100.00%	9.00
P2	20.045	90.91%	18.333	100.00%	19.351	94.59%	8.00
P3	13.133	93.33%	5.467	93.33%	9.300	93.33%	2.00
P4	27.000	100.00%	12.000	93.33%	19.500	96.67%	10.00
P5	14.276	93.10%	11.533	100.00%	13.341	95.45%	9.00
P6	51.800	86.67%	6.333	100.00%	29.067	93.33%	9.00
P7	26.417	100.00%	8.800	100.00%	19.641	100.00%	7.00
P8	21.467	100.00%	6.467	93.33%	13.967	96.67%	8.00
P9	15.680	96.00%	13.600	100.00%	14.900	97.50%	10.00
P10	25.238	95.24%	11.867	100.00%	19.667	97.22%	9.00
Average	23.506	95.53%	10.147	98.00%	17.319	96.48%	8.10

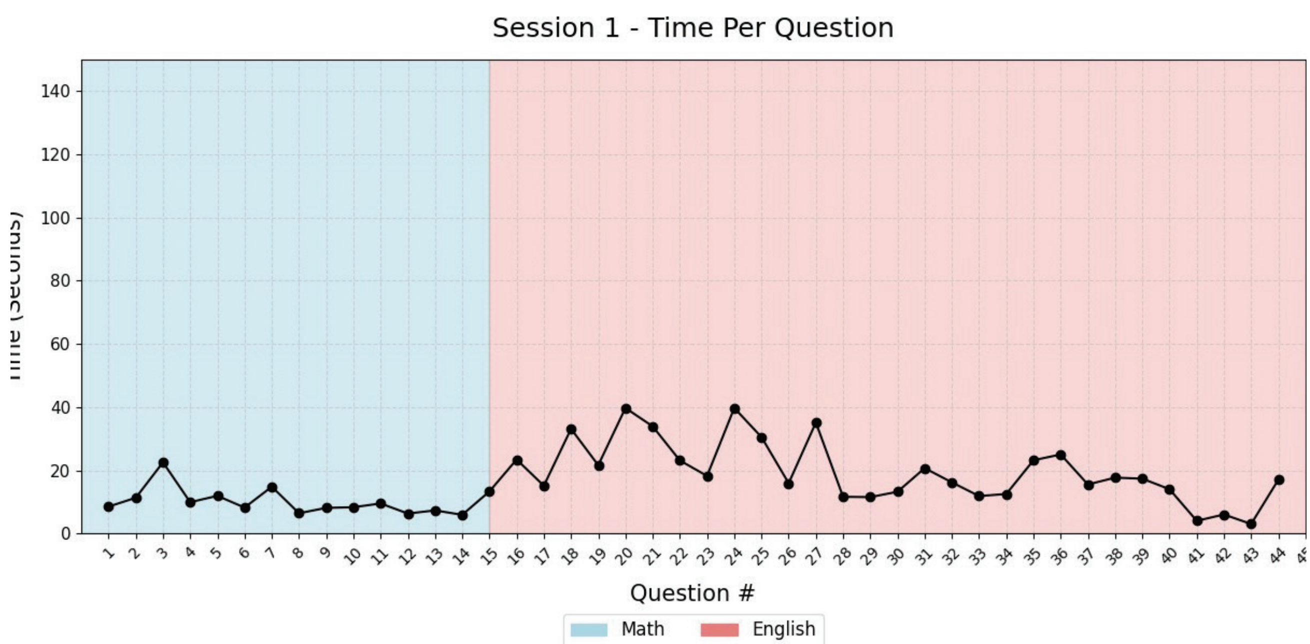

Figure 1. *Average response time per question during Session 1. Mean response time across 10 participants during the uninterrupted baseline session. The x-axis shows question order, and the y-axis shows average response time in seconds. The shaded regions distinguish mathematics and English questions.*

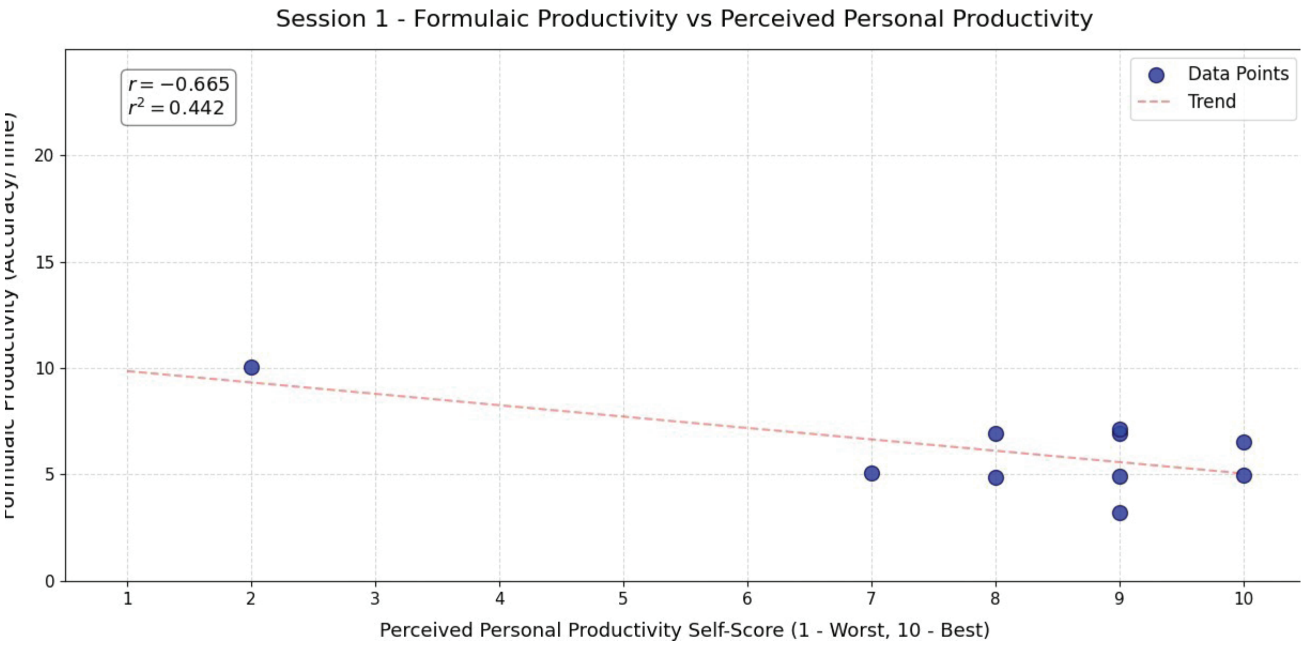

Figure 2. *Relationship between perceived and formula-based productivity during Session 1. Scatterplot comparing participants' Perceived Personal Productivity (PPP) scores with formula-based productivity scores during the uninterrupted session. Each point represents one participant. The fitted trendline and correlation statistics summarize the association between the two measures.*

Table 4. *Consolidated Data Collected from Session Two Per Participant.*

Participant	ET	EA	MT	MA	AT	TA	PPP
P1	36.042	100.00%	9.200	100.00%	25.718	100.00%	8.00
P2	20.000	100.00%	14.000	93.33%	17.000	96.67%	9.00
P3	14.133	100.00%	5.200	100.00%	9.667	100.00%	1.00
P4	13.333	100.00%	8.467	93.33%	10.900	96.67%	10.00
P5	14.923	100.00%	16.467	93.33%	15.488	97.56%	7.00
P6	59.933	100.00%	31.333	93.33%	45.633	96.67%	6.00
P7	16.286	100.00%	11.733	100.00%	14.698	100.00%	8.00
P8	12.867	100.00%	5.067	100.00%	8.967	100.00%	9.00
P9	16.667	100.00%	11.867	100.00%	14.267	100.00%	8.00
P10	20.385	100.00%	9.667	100.00%	16.463	100.00%	7.00
Average	22.457	100.00%	12.300	97.33%	17.880	98.76%	7.30

Note. "ET" refers to Average English Time Per Question, "EA" refers to Total English Accuracy, "MT" refers to Average Math Time Per Question, "MA" refers to Total Math Accuracy, "AT" refers to Average Time Per All Questions, and "TA" refers to Total Accuracy. 'P1' through 'P10' denote Participants 1 through 10.

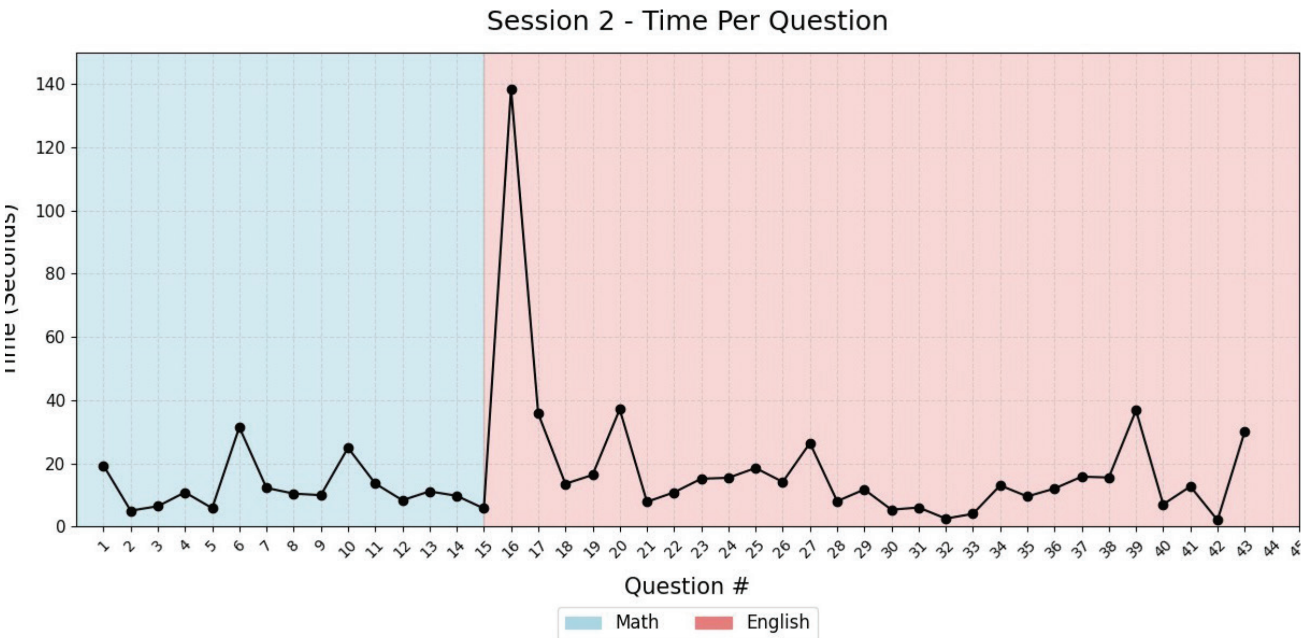

Figure 3. *Average response time per question during Session 2. Mean response time across 10 participants during the session containing a planned one-minute break between the mathematics and English sections. The x-axis shows question order, and the y-axis shows average response time in seconds. The prominent increase after the break indicates delayed resumption of task performance.*

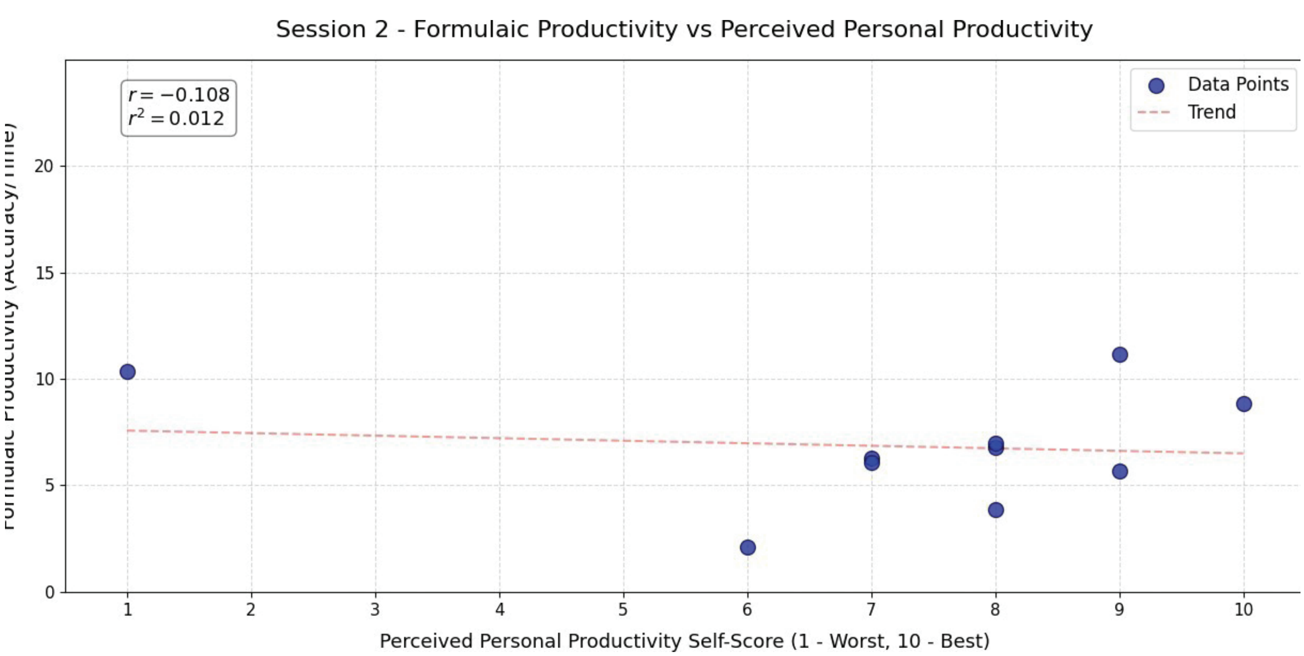

Figure 4. *Relationship between perceived and formula-based productivity during Session 2. Scatterplot comparing participants' PPP scores with formula-based productivity scores during the planned-break session. Each point represents one participant. The fitted trendline and correlation statistics summarize the association between subjective and calculated productivity.*

Session Three

The caveat for Session three was that at a random point throughout the session, each participant was given a mentally stimulating riddle that they had to solve before moving on to their next question. The results, separated by each of the participants, are summarized in Table 5.

The data from session three revealed very similar outcomes, but this time with more variance among the participants themselves (Table 5). The biggest change seen in this session from the previous two sessions was that P6, who had both the lowest time and accuracy outcomes, had the highest accuracy this session, and a relatively lower time compared to the other participants. Along with a big performance jump for P6, they gave themself a perfect PPP score. The consistency of participants' time per question during this session is illustrated in Figure 5, and the relationship between participants' self-reported PPP scores and their calculated formulaic productivity scores is shown in Figure 6.

DISCUSSION

Analysis of Results

The findings from the study, in both pure data and graphical forms, were analyzed to find possible interpretations.

Session One

In session one, the consistency in participants' time per question–as seen in Figure 1–possibly indicated that participants could have had a constant state of workflow during session 1, which aligns with the framework proposed by Kim *et al*., who identified consistency of effort as a component of efficiency in personal productivity (Figure 1) (2). However, in regards to the participants' reported PPP scores, the pattern of overestimation–as seen in Figure 2–suggested that the simple and consistent state of workflow that participants experienced during session one could have led them to believe that they were more productive in this span than they actually were (Figure 2). This also suggested a potential flaw in the basic form of formulaic productivity, as it doesn't account for continuity and consistency. This potential flaw was important to account for during the construction of the mathematical model.

Session Two

In session two the large spike in time per question–as seen in Figure 3–suggested that participants may have taken a long time to adjust back to the mindset of answering questions because of the lull period that existed during their one-minute break (Figure 3). However, after the first question, the participants may have been able to

Table 5. *Consolidated Data Collected from Session Three Per Participant.*

Participant	ET	EA	MT	MA	AT	TA	PPP
P1	33.474	100.00%	17.533	93.33%	26.441	97.06%	8.00
P2	16.158	100.00%	25.133	80.00%	20.118	91.18%	7.00
P3	2.400	100.00%	6.800	93.33%	4.600	96.67%	2.00
P4	9.067	100.00%	10.200	100.00%	9.633	100.00%	7.00
P5	5.143	100.00%	30.533	93.33%	18.276	96.55%	5.00
P6	11.933	100.00%	8.667	100.00%	10.300	100.00%	10.00
P7	1.133	100.00%	22.933	93.33%	12.033	96.67%	5.00
P8	8.400	100.00%	9.467	100.00%	8.933	100.00%	6.00
P9	1.400	100.00%	16.067	100.00%	8.733	100.00%	7.00
P10	18.320	100.00%	13.267	86.67%	16.425	95.00%	8.00
Average	10.743	100.00%	16.060	94.00%	13.549	97.31%	6.50

Note. "ET" refers to Average English Time Per Question, "EA" refers to Total English Accuracy, "MT" refers to Average Math Time Per Question, "MA" refers to Total Math Accuracy, "AT" refers to Average Time Per All Questions, and "TA" refers to Total Accuracy. 'P1' through 'P10' denote Participants 1 through 10.

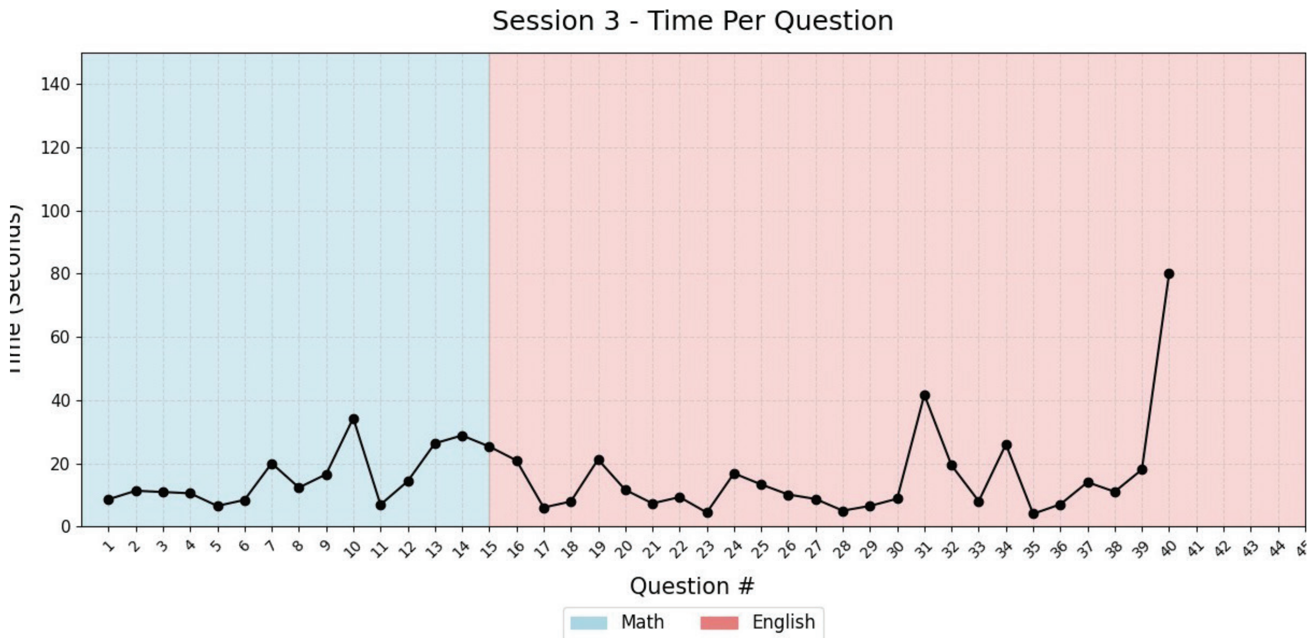

Figure 5. *Average response time per question during Session 3. Mean response time across 10 participants during the sudden-interruption session. Participants received a mentally engaging riddle at a randomly assigned point during the assessment. The x-axis shows question order, and the y-axis shows average response time in seconds. The shaded regions distinguish mathematics and English questions.*

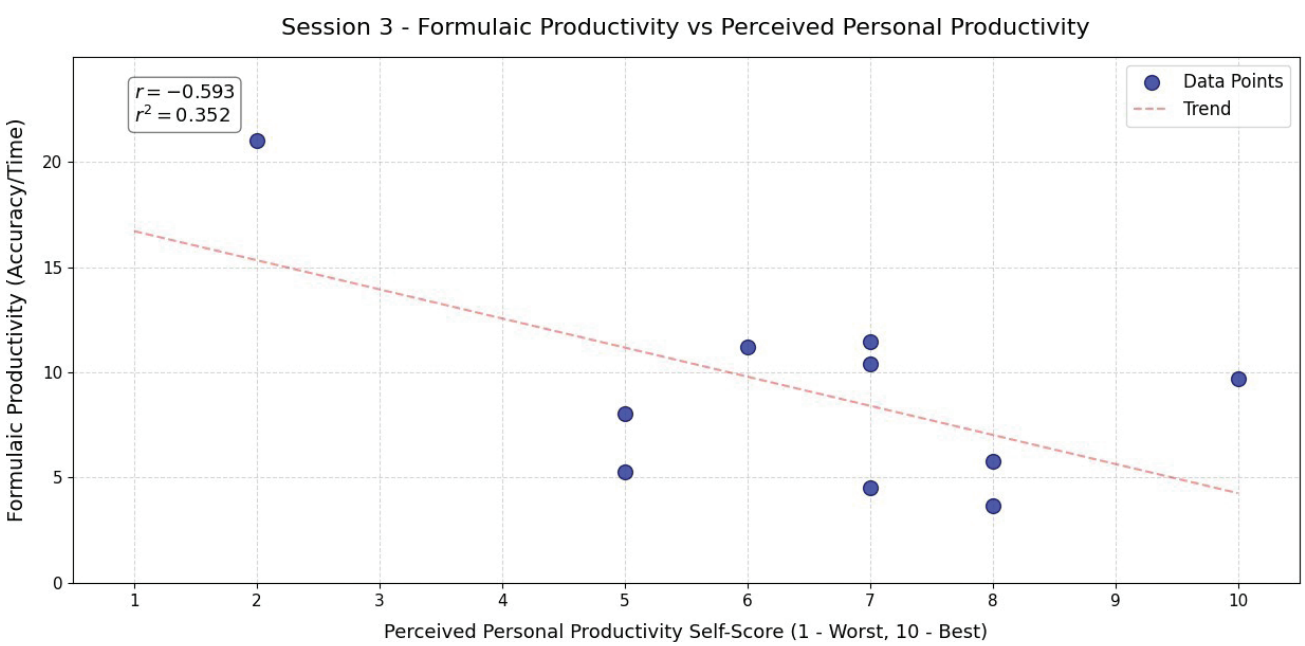

Figure 6. *Relationship between perceived and formula-based productivity during Session 3. Scatterplot comparing participants' PPP scores with formula-based productivity scores during the sudden-interruption session. Each point represents one participant. The fitted trendline and correlation statistics summarize the association between subjective and calculated productivity. Note. Correlational analysis between self-reported PPP scores and mathematically calculated formulaic productivity scores for Session 3, based on an adapted version of the outputs/inputs industry productivity formula. PPP is measured on a modified 1-10 SPS (y-axis), while formulaic productivity (x-axis) evaluates output metrics derived by dividing the accuracy by the time taken to answer the questions. The trendline indicates a near inverse relationship between participants' self-reported PPP scores, and the base formulaic productivity output.*

reenter their state of workflow for the rest of the test, as suggested by the more consistent level of productivity not only before, but after the spike as well. This pattern aligns with the findings of Yi and colleagues, who demonstrated that in manufacturing environments with repetitive tasks, productivity declines immediately after planned breaks due to disrupted work rhythm and reduced working proficiency (4). Interestingly, the high level of inaccuracy of the participants' reported PPP scores–as seen in Figure 4–likely suggested that session two possibly had a disorienting effect on the participants due to the planned break, which could have led them to not have a proper idea on how productive they actually were (Figure 4). This possible association between disorientations and inaccuracy in PPP highlighted a potential flaw in the usage of PPP to calculate personal productivity.

<u>Session Three</u>

The fluctuations in participants' average time per question during session three–as seen in Figure 5–can likely be attributed to two main factors: participants' reaction to riddles, and the design of the assessment website (Figure 5). Firstly, similar to how participants struggled to adjust after the one-minute break from session two, participants may have felt distracted by the riddle's sudden interruption, which could have led them to lose focus for a short period of time after their riddle. This finding is consistent with Borst and colleagues' problem-state bottleneck theory, which showed that interruptions requiring active mental processing (such as the riddles in Session 3) disrupt task resumption more severely than interruptions that do not demand such cognitive engagement (1). Secondly, the website's design had a flaw in that it showed participants the riddle at a purely random time, meaning that each participant's potential loss of focus would have occurred at a different period of time during the session. Because of this, the spikes in Figure 5 are smaller, as most of the time, only one participant received a riddle at that question (Figure 5). However, the exception to this came at the end of the session, as if by the last question, if the participant had not received a riddle, they were given one. Thus, this safety measure resulted in more than one participant receiving a riddle before their last question which is consistent with the largest spike in Figure 5 existing at question 40, the last question (Figure 5). The inverse relationship between session three's formulaic productivity and PPP scores–as seen in Figure 6–likely indicated that the riddles of session three may have not only made the participants incorrectly perceive their productivity, but it could have also made them think they performed completely opposite to their actual performance (Figure 6). However, this highlighted a potential limitation of the system of calculating productivity through accuracy divided by time–an applied form of outputs divided by inputs–as if a student spends a large amount of time to g*et al*l questions answered, then their accuracy doesn't increase by a large margin, but their time does, which would cause the formulaic productivity value to decrease.

Model Conceptualization

The experiment mainly assessed three of Kim's defined four main aspects of personal productivity: amount (total questions attempted); accuracy (proportion correct); efficiency, the relationship between quantity and time; and feeling, the self perception of personal productivity (2). While Gennara validated the SPS as a reliable measure of perceived effectiveness and efficiency, this study suggested that self-reported PPP exhibited weak correlation–and in some cases inversely related–with objective performance metrics (5). This discrepancy may be due to the nature of the assessments used in this study compared to the study from Gennara, which focused on the broader workplace context (5). Thus, the PPP metric was not factored into the final model. However, Kim's remaining two aspects: amount and efficiency, were indeed factored into the final model, as from the results of the experiment, these metrics appeared to be useful indicators of participant productivity, though this finding requires validation in larger independent samples (2). Continuing to draw from the results, the breaks and interruptions appeared to influence both of the aforementioned metrics of productivity, and therefore, they may have also been factors in the final model. However, there was not a differentiation between interruptions and breaks, because per the study, they both had similar effects on productivity, especially if the interruptions from session three were given at the same random time for all participants. As such, for the purposes of this model, interruptions and breaks were both treated as breaks.

Variable Inputs

The amount component of productivity was represented by Q, accuracy was represented by TA (a proportion from 0 to 1), efficiency was evaluated as (Q/TT), and the factor of breaks was integrated as proportion of the total time that breaks covered (T_{break}/TT). The model ultimately outputs a personal productivity score (P).

Exponential and Multiplicative Weights

To assign certain weightage to each of the factors: accuracy, efficiency, and break time, weights of alpha (α), beta (β), and delta (δ), respectively, were assigned. The alpha and beta weights are both exponential weights, as they serve to increase the leverage of both accuracy and efficiency as those values get higher, due to their importance. The delta weight, however, is multiplicative, as it is intended to have a constant effect despite a change in the proportion of break time. This decision was made in accordance with the results of the conducted experiment, as the severity of the break resulted in negligible change in productivity effect, and therefore, the leverage of breaks doesn't need to become increasingly important as it increases in value. The maximum value for all the weights is 2, and the minimum value is 0. The upper bound was chosen so that at most, the alpha and beta values individually increase the importance of total accuracy and efficiency by a quadratic factor, and the delta weight would double the value of break time. Thus the lower bound of 0 for these values serves to nullify the component of the formula which the weight is being applied to.

Normalization of Values

By relating the listed variables in a general division statement, the resultant values are arbitrary, having no defined scale. To resolve this issue, the raw productivity score was normalized by subtracting the minimum productivity score (P_{min}). After this reduction, the remaining value was divided by the range of productivity scores (P_{max} - P_{min}). This resulted in the productivity model producing values on a normalized scale from 0 to 1, so the results were systematic. However, this form of the model was also unhelpful, as neither P_{max} nor P_{min} were numerically defined. The value of P_{min} was relatively simpler to establish, as during a span of time, the lowest productivity one can express is zero, requiring no progress whatsoever to be made.

In order to ascertain the values of P_{max} , however; the model was evaluated at its theoretical maximum: Total Accuracy (TA) = 1, Quantity (Q) = 100 (the maximum quantity set for this model due for practical constraints), Total Time (TT) = 1 minute (the minimum meaningful work duration set for this model), and Break Time (T_{break}) = 0. With baseline weights ($\alpha = \beta = \delta = 1$) (α=β=δ=1), this yields a max productivity score of 1, as the numerator and denominator both equate to 100. This is the reason for the factor of 100 in the denominator of the model equation–it ensures that the maximum possible unscaled score is exactly 1 at the theoretical maximum inputs. Thus, on the unscaled (0-1) metric, P_{max} = 1. When multiplied by 100 for the final interpretable score, the maximum possible score is 100. It is important to note that this derivation assumes the baseline weights ($\alpha = \beta = \delta = 1$). If the weights are altered, the maximum score may change, and the normalization should be recalculated accordingly. Once the maximum and minimum values for Productivity were determined, they were entered into the normalized formula, which gave an unsimplified, yet normalized productivity formula which outputs values on a scale from 0 to 1.

Constructed Model

The final step in the construction of the productivity mathematical model, after the normalization process, was to make the values easily interpretable. The normalization process described above produces values on a normalized scale from 0 to 1. To make the values more interpretable for high school students, these values can be multiplied by a factor of 100, producing a score on a scale from 0 to 100 (Figure 7). It is important to note that the validation tests reported in Figure 8 are based on the unscaled (0-1) values, prior to multiplication by 100.

$$P = \frac{TA^{\alpha}\left(\frac{Q}{TT}\right)^{\beta}}{100\left(1+\delta\left(\frac{T_{\text{break}}}{TT}\right)\right)}$$

Figure 7. *Proposed mathematical model for personal productivity. The model estimates personal productivity using total accuracy, quantity of completed questions, total time, break time, and adjustable weighting parameters. P denotes personal productivity; TA, total accuracy expressed as a proportion from 0 to 1; Q, total questions attempted; TT, total time in minutes; T_break, break time in minutes; and α, β, and δ, weighting parameters for accuracy, efficiency, and break effects, respectively. The unscaled output ranges from 0 to 1 and may be multiplied by 100 to produce a 0–100 score under the specified baseline normalization assumptions.*

Model Validity

It was not sufficient to create the model, for it was important to ensure that the model was valid and reliable. In order to assess the validity, five tests were performed, re-utilizing the collected data from the experiment as test data (Figure 8).

```
======================================================================
TEST 1: Mean Score Comparison Across Sessions
======================================================================

  Session                               Mean Score
  -----------------------------------  ------------
  Session 1 (No breaks)                   0.026494
  Session 2 (1-min break)                 0.026782
  Session 3 (Sudden Interruptions)        0.029250

  S1 → S2 change : +1.09%
  S2 → S3 change : +9.21%
  S1 → S3 change : +10.40%

 Monotone decrease (S1 ≥ S2 ≥ S3): False

======================================================================
TEST 2: Correlation of Model Score vs Perceived Productivity (PPP)
======================================================================

  Session                              Pearson r     p-value   Spearman ρ     p-value
  -----------------------------------  ----------  ----------  -----------  ----------
  Session 1 (No breaks)                  -0.6627      0.0263      -0.2431      0.4714
  Session 2 (1-min break)                -0.0996      0.7707       0.2765      0.4104
  Session 3 (Sudden Interruptions)       -0.5430      0.0843      -0.3410      0.3047
  All sessions (pooled)                  -0.4194      0.0151      -0.1195      0.5076

 Pooled Pearson significance: ✓ significant (p < 0.05)

======================================================================
TEST 3: Parameter Sensitivity Analysis (using All-Session data)
======================================================================

  Baseline (α=1, β=1, δ=1) mean score : 0.026330

  Parameter     Value   Mean Score   Rank Corr w/ Baseline    p-value
  ------------  ------  ------------  ----------------------  ----------
  alpha           0.5     0.026654                  0.9909      0.0000
  alpha           1.0     0.026330                  1.0000      0.0000 ← baseline
  alpha           1.5     0.026010                  1.0000      0.0000
  alpha           2.0     0.025696                  1.0000      0.0000
  beta            0.5     0.015289                  1.0000      0.0000
  beta            1.0     0.026330                  1.0000      0.0000 ← baseline
  beta            1.5     0.046663                  0.9909      0.0000
  beta            2.0     0.085127                  0.9909      0.0000
  delta           0.5     0.027294                  1.0000      0.0000
  delta           1.0     0.026330                  1.0000      0.0000 ← baseline
  delta           1.5     0.025437                  1.0000      0.0000
  delta           2.0     0.024609                  1.0000      0.0000
```

Figure 8. *Summary of model evaluation analyses. Python-based evaluation of the proposed productivity model using data from 10 participants across three assessment sessions. The figure reports mean score comparisons, correlations between model scores and PPP, parameter sensitivity, break-penalty consistency, and rank stability across sessions. All model scores are shown on the unscaled 0–1 metric, and statistical significance was assessed using two-tailed tests with $p < 0.05$.*

Full Validity Test Output Cont.

TEST 4: Break Penalty Validity (per-participant)

Participant	Score (Session 1)	Score (Session 2)	Score (Session 3)	Consistent?
Participant 1	0.018810	0.018003	0.017262	✓
Participant 2	0.021056	0.019998	0.019042	✓
Participant 3	0.052092	0.046210	0.041522	✓
Participant 4	0.031020	0.028858	0.026977	✓
Participant 5	0.026580	0.024956	0.023519	✓
Participant 6	0.014444	0.013951	0.013491	✓
Participant 7	0.026715	0.025116	0.023698	✓
Participant 8	0.039412	0.036021	0.033167	✓
Participant 9	0.032589	0.030238	0.028203	✓
Participant 10	0.023658	0.022374	0.021223	✓
Participant Avg.	0.025369	0.023901	0.022593	✓

Participants where penalty is monotone (✓): 11/11

TEST 5: Rank Stability Across Sessions

Participant	Rank S1	Rank S2	Rank S3	Rank Var
Participant 1	1	2	1	0.33
Participant 2	2	6	9	12.33
Participant 3	3	1	3	1.33
Participant 4	4	10	11	14.33
Participant 5	5	4	2	2.33
Participant 6	6	9	7	2.33
Participant 7	7	5	6	1.00
Participant 8	8	3	4	7.00
Participant 9	9	7	8	1.00
Participant 10	10	8	10	1.33
Participant Avg.	11	11	5	12.00

```
Rank correlation S1 vs S2 : ρ = 0.5273  (p = 0.0956)
Rank correlation S1 vs S3 : ρ = 0.2636  (p = 0.4334)
Rank correlation S2 vs S3 : ρ = 0.7000  (p = 0.0165)
Average cross-session rank correlation : ρ = 0.4970
```

Continued Figure 8. *Summary of model evaluation analyses. Python-based evaluation of the proposed productivity model using data from 10 participants across three assessment sessions. The figure reports mean score comparisons, correlations between model scores and PPP, parameter sensitivity, break-penalty consistency, and rank stability across sessions. All model scores are shown on the unscaled 0–1 metric, and statistical significance was assessed using two-tailed tests with $p < 0.05$.*

Mean Score Comparison

The goal of the Mean Score Comparison test was to cross-reference the trend of productivity from the experiment (which outputted a decreasing productivity score for each session) with the average score of the participants outputted by the model from the sessions (Figure 8). The output of this validity test indicated that the direct comparison of mean scores didn't reflect the trend from the experiment.

PPP Correlation

The PPP Correlation test assessed discriminant validity–specifically, whether the model's objective productivity scores diverged from subjective self-perceptions (PPP). This test examined whether the model output aligns with the trend in PPP self-reported scores from the participants of the experiment (The trend showed decreasing PPP each session). However, the outcome of this test showed that the model scores exhibited a divergent pattern relative to the PPP values, which is consistent with the hypothesis that objective productivity measures may diverge from subjective perceptions.

Parameter Sensitivity Analysis

Parameter Sensitivity Analysis assessed construct validity–specifically, whether the model's outputs remain stable across reasonable variations in its parameters. This test compared the outputted model score depending on the values of each of the weights, alpha (α), beta (β), and delta (δ), and checked for a strong correlation between the scores (Figure 8). The output of the Parameter Sensitivity Analysis indicated an extremely strong correlation between the mean scores across the variant values of the weights, suggesting that the model demonstrates preliminary support for this validity criteria.

Break Penalty Validity

The Break Penalty Validity test assessed internal consistency–specifically, whether the break penalty factor (T_break/TT) was applied uniformly across all participants. This test served to isolate the factor regarding the proportion of break time to ensure that the penalty received from the breaks was consistent for every participant (Figure 8). The output of the Break Penalty Validity Test indicated a consistent output from the factor of the model regarding proportion of break time, with all ten participants showing consistency across all three sessions. This provides preliminary evidence supporting the model's validity in this regard.

Rank Stability Analysis

Rank Stability Analysis assessed reliability–specifically, whether the model produces consistent relative rankings of participants across different session conditions. This test compared the outputted model score for each of the participants across the sessions, and checked the consistency of the participants' rank amongst themselves (Figure 8). The output of the Rank Stability Analysis indicated a correlation of moderate strength in the ranks of the participants across sessions based on the outputted model score, suggesting that the model demonstrates some stability in relative rankings, though further testing with larger samples would strengthen this finding.

CONCLUSION

The goal of this paper was to construct a mathematical model which could quantify personal productivity for high school students by analyzing the relationship between a myriad of factors which interact with productivity, and understand the importance and weightage of different relationships.

One potential implication of this research is that the preliminary findings and the proposed model could serve as a framework for high school students to estimate their day-to-day productivity if they input the necessary variables and weights. Unlike previous applications used to track productivity by logging activities, such as the OmniTrack system proposed by Kim *et al*. or the application proposed by Romanov, this model may allow students to obtain a quantitative score that informs them of their estimated productivity each day, rather than relying solely on subjective estimates based on logged activities (6, 12).

With further validation, this approach could also enable students to identify patterns in their productivity through a feedback loop, potentially supporting efforts to improve their work habits. This method of productivity tracking could also be adapted for other settings, allowing more people to potentially manage their personal productivity. While the study focuses on students, many of the components assessed may generalize to other populations, though this requires further investigation.

There were some sampling restrictions that existed in the experiment, with the sample size of 10 participants not passing the requirement for the Central Limit Theorem needed to draw conclusions from mean-based data. Though the requirement to assume normality and therefore draw conclusions from means was met by plotting the data and checking for outliers, it would be more ideal to have a large enough size, with over 30 participants, to increase the statistical power of the test. Additionally, while participants were asked to respond as honestly as possible, there was no sound method of verifying that participants followed these instructions, leaving room for potential bias in the response. While the constructed model is relatively flexible in regards to other Human-Computer Interaction based mathematical models, compared to other applications that could be used to track metrics, it still lacks a level of customization or flexibility to enter in assignment-specific metrics, and can only be used in a general sense that overlaps most disciplines. To improve the adaptability, future research is recommended to venture in a computational-focused direction, as it would allow for a more robust and flexible model. Finally, in the future, it is recommended that research is done that compares the three mentioned aspects of productivity, as well as combining the 4th aspect: alignment, into the study to do a more holistic analysis of productivity.

The findings from the research suggested three main ideas. Firstly, the relationship between quantitative aspects of productivity and the perception of one's productivity appeared to be inconsistent within the confines of this experiment, as mental aspects of assessments may have impacted perception of productivity too heavily for it to serve as a standalone metric. Secondly, the main values that impacted student productivity in this experiment were amount as well as efficiency. Finally, the results of the study serve as a possible example as to how, despite technological and theoretical advancements, the construction of a model that accurately quantifies productivity across all scenarios was still difficult.

CONFLICT OF INTEREST

The author declares that there are no conflicts of interest related to this work.

REFERENCES

1. Borst JP, Taatgen NA, Van Rijn H. What makes interruptions disruptive? A process-model account of the effects of the problem state bottleneck on task interruption and resumption. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15). *Seoul (KR): ACM*. 2015; p. 2972-2976. https://doi.org/10.1145/2702123.2702156
2. Kim Y, Choe EK, Lee B, Seo J. Understanding personal productivity: How knowledge workers define, evaluate, and reflect on their productivity. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). *Glasgow (UK): ACM*. 2019; p. 1-12. https://doi.org/10.1145/3290605.3300845
3. Günter A, Gopp E. Overview and classification of approaches to productivity measurement. *Int J Product Perform Manag.* 2022; 71 (4): 1212-1229. https://doi.org/10.1108/IJPPM-05-2019-0241
4. Yi S, Gong Q, Dong F, Wang H. The effect of planned breaks on worker productivity and the moderate role of workload in a manufacturing environment. *Asian Econ Financ Rev.* 2020; 10 (12): 1366-1383. https://doi.org/10.18488/journal.aefr.2020.1012.1366.1383
5. Gennara AK. Understanding and measuring perceived productivity (master's thesis). Ottawa (ON): Carleton University; 2023.
6. Kim Y, Jeon JH, Lee B, Choe EK, Seo J. OmniTrack: A flexible personal tracking system for everyday life. *Proc ACM Interact Mob Wearable Ubiquitous Technol*. 2017; 1 (3): 1-28. https://doi.org/10.1145/3130928
7. Pedersen M, Muhr SL, Dunne S. Time management between the personalisation and collectivisation of productivity: The case of adopting the Pomodoro time-management tool in a four-day workweek company. *Time Soc*. 2024; 33 (4): 417-437. https://doi.org/10.1177/0961463X241258303
8. Ko AJ. Why we should not measure productivity. In: Apress eBooks. New York: Apress. 2019; p. 21-26. https://doi.org/10.1007/978-1-4842-4221-6_3
9. Maniktala M, Chi M, Barnes T. Enhancing a student productivity model for adaptive problem-solving assistance. *User Model User-Adapt Interact*. 2022; 33 (1): 159-188. https://doi.org/10.1007/s11257-022-09338-7
10. Mazur JE. Mathematical models and the experimental analysis of behavior. *J Exp Anal Behav*. 2006; 85 (2): 275-291. https://doi.org/10.1901/jeab.2006.65-05
11. Vasileiou K, Barnett J, Thorpe S, Young T. Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Med Res Methodol*. 2018; 18 (1): 148. doi: 10.1186/s12874-018-0594-7. https://doi.org/10.1186/s12874-018-0594-7
12. Romanov P. Application for personal productivity tracking (Internet) (bachelor's thesis). Vytautas Magnus University: (publisher unknown); 2025 (cited 2026 May 18). Available from: https://portalcris.vdu.lt/server/api/core/bitstreams/a71b67f0-b0f3-4c79-beb9-ea87518f1920/content