







The elusive relationship between time on-task and learning: not simply an issue of measurement

Karrie E. Godwin^{a,b} , Howard Seltman^c , Ma Almeda^d, Mandi Davis Skerbetz^{e,f,g}, Shimin Kai^d, Ryan S. Baker^h  and Anna V. Fisherⁱ 

^aDepartment of Psychology & the Sherman Center for Early Learning in Urban Communities, University of Maryland Baltimore County, Baltimore, MD, USA; ^bDepartment of Educational Psychology, Kent State University, Kent, OH, USA; ^cDepartment of Statistics, Carnegie Mellon University, Pittsburgh, PA, USA; ^dDepartment of Human Development, Teachers College Columbia, New York, NY, USA; ^eJohns Hopkins University, Baltimore, MD, USA; ^fDallas Academy, Dallas, TX, USA; ^gSouth Fayette Township School District, McDonald, PA, USA; ^hUniversity of Pennsylvania, Graduate School of Education, Philadelphia, PA, USA; ⁱDepartment of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

According to the time on-task hypothesis, the amount of time an individual devotes to an instructional task determines the extent to which learning occurs. Therefore, time off-task hampers learning by limiting learning opportunities. Prior research has generally found a positive relationship between time on-task and achievement; however, the correlation strength is highly variable across studies. Differences in the ways in which time has been operationalised may be one factor contributing to the divergent results. We utilise an existing data set of twenty classrooms (K-4) to investigate whether operationalising time in a consistent manner will yield a stable association between on-task behaviour and learning. Overall, on-task behaviour was positively correlated with learning, controlling for gender, school type, and grade-level. However, this correlation was weak. Importantly, considerable variability in the correlation strength was observed, indicating variations in the prior literature cannot be attributed solely to issues of measurement.

ARTICLE HISTORY

Received 13 February 2020
Accepted 19 February 2021

KEYWORDS

On-task behaviour; off-task behaviour; attention; learning; achievement

Introduction

A common belief in education is that better learning outcomes should result the more time students spend on a given task. This popular assumption was characterised by Carroll (1963) as the *Time-on-Task hypothesis*. The *Time-on-Task hypothesis* prompted an up swell of research aiming to demonstrate that learning is a function of the amount of time spent on a specific task (Cobb, 1972; Fredrick et al., 1979; Lahaderne, 1968; McKinney et al., 1975; Samuels & Turnure, 1974). Practitioners and policy makers

CONTACT Karrie E. Godwin  kgodwin@umbc.edu  Department of Psychology & the Sherman Center for Early Learning in Urban Communities, University of Maryland Baltimore County, 1000 Hilltop Cir, Baltimore, MD 21250, USA

are also interested in research on time as they are committed to identifying predictors of achievement that are malleable. Unfortunately, many of the strongest predictors of achievement are not amenable to intervention (see Karweit & Slavin, 1980 for discussion). However, time is (in principle) malleable and therefore it could be targeted through interventions that aim to increase instructional time or alternatively optimise how instructional time is used.

Despite the extensive body of research investigating this topic, the relationship between time and learning remains elusive as prior research has obtained mixed findings. In the prior literature, a positive association between time spent on-task and achievement has typically been found, yet the correlation strength vacillates *dramatically* across studies (e.g. see Karweit, 1984 for review; for more contemporary examination of this topic see: Baker, Corbett, et al., 2004; Connor et al., 2009; Fisher et al., 2014; Gobel, 2008; Godwin, 2015; Godwin & Fisher, 2014; Kovanovic et al., 2015; Kupiainen et al., 2014; Moffett & Morrison, 2020; Roberge et al., 2012). Indeed, in our review we found estimates for the correlation strength between measures of time and learning/performance ranged between $-.23$ and $.78$ (similar ranges were found for the association between measures of *time off-task* and learning/performance with estimates ranging from $.07$ to $-.53$; although note that some significant positive estimates have also been obtained; e.g. Choudhury and Gorman (2000). See Table 1 for a selective overview of the correlations between learning outcomes and different measurement levels of time (i.e. student level, classroom level, and quantity of schooling) for a variety of grade levels from kindergarten to high school. For similar discussion regarding the relationship between homework time and achievement see Cooper et al. (2006) and Flunger et al. (2015).

The reasons for the wide variability in the estimates of the strength of the relationship between time on-task and learning are not fully understood; however, recent research may point to an explanation for this phenomenon. It is possible not all tasks are equally beneficial for learning and thus the relationship between time and learning may not be uniform across tasks. In line with the time-on-task hypothesis, Carvalho et al. (2017) found time on-task was a significant predictor of adult students' quiz and exam scores: students were more likely to obtain higher scores the more time they spent reading assignments and completing instructional activities. Critically, time spent on these different tasks did not translate into equal learning benefits. On average students needed to spend 13.8 h per week reading compared to 1.5 h per week completing activities to obtain the same 1 *SD* increase in quiz scores. This finding was taken as support for the 'doer effect', but simultaneously provides evidence the relationship between time and learning is not straightforward and the magnitude of the hypothesised benefits of time may not be due solely to the quantity of time spent on-task but may vary based on the quality and nature of the learning activities.

Karweit and Slavin (1982) provide another feasible explanation for the mixed findings; across the prior literature, in addition to methodological and procedural differences, researchers have utilised numerous operational definitions of *time* (see Caldwell et al., 1982; Fredrick & Walberg, 1980; Goodman, 1990; Karweit & Slavin, 1981; Wiley & Harnischfeger, 1974). For instance, time has been defined at the *student-level* and thus operationalised as on-task behaviour, engaged time, and looking time—proxies for

Table 1. Correlations between time and learning outcomes: a selective summary of the prior literature.

Author (year)	Grade level	Level of time measure	Correlation between learning/ performance and time on-task	Correlation between learning/ performance and time off-task
Baker, Corbett, et al. (2004)	Middle School	Student level	Mean positive correlations: .09 to .30 Mean negative correlations: -.10 to -.23	-.08 to -.38*
Baker et al. (2004)	Middle/High School	School quantity level and Classroom level		
Cobb (1972)	Elementary	Student level	.24 to .48 ^a	-.11 to -.43 ^a
Edmiston and Rhoades (1959)	High School	Student level	.51 to .58 ^a	
Everston et al. (1980)	Junior high	Student level	.29* to .39*	-.14 to -.33* -.50*
Fisher et al. (2014)	Kindergarten	Student level		
Fredrick et al. (1979)	High School	Hybrid: School quantity/ Classroom/Student Level	.54*	
Godwin and Fisher (2014)	Kindergarten	Student level		-.38 ⁺
Karweit and Slavin (1981)	Elementary	Classroom level and Student level	.189 to .422*	
Lahaderne (1968)	Middle School	Student level	.37* to .53*	-.38* to -.53*
Lee et al. (1999)	K-12	Student level	Work completion: .43* Quality: .004	
Moffett and Morrison (2020)	Kindergarten	Student level		Achievement in 1st grade: .07 to -.17*
Odell (1923)	Elementary/Middle School	School quantity level	.08 to .14 ^a	-.0229*
Roberge et al. (2012)	Middle School	Student level		
Roby (2003)	Elementary, Middle, & High School	School quantity level	.54* to .78*	
Rozelle (1968)	High School	School quantity level	.046 to .126 ^a	.13 to .36 ^a
Smith (1979)	Elementary	School quantity level, Classroom level, and Student level	.13 to .36 ^a	

Note. The table provides the correlation coefficients between measures of time and learning/performance outcomes by study as well as contextual information including the grade level of the participants and the level of measurement for time (i.e. student level measures which include proxies for attention such as engaged time, looking time, or proportion of on/off-task behaviour; classroom level measures such as the amount of time scheduled for instruction, as well as school quantity including length of the school day or year and attendance). Note that some studies report measures of time *off-task* (see column 5) as opposed to *on-task* (see column 4) these values are noted in separate columns to aid interpretation. Unless otherwise specified, multiple values reflect the range of correlations reported in a study.

Significant correlations ($p < .05$) are denoted by *, + indicates marginally significant correlations ($p < .10$); “a” indicates that significance was not reported.

attention (Choudhury & Gorman, 2000; Cobb, 1972; Fisher et al., 2014; Godwin, 2015; Godwin & Fisher, 2014; Godwin et al., 2016; Lahaderne, 1968; Lee et al., 1999). Time has also been defined at the *classroom-level*, and thus operationalised as the amount of time allotted or the time spent on a specific subject area or activity (Arlin & Roth, 1978; Baker et al., 2004). Still others have connected time to the quantity of schooling students' receive and operationalised time as the length of the school day, length of the school year, total number of school days attended, or years of schooling (Agrawal, Smith, & Wick, 1977 as cited in Fredrick & Walberg, 1980; Cooper et al., 2010; Hough & Bryde, 1996; Hyman et al., 1975; Karweit, 1973; Roby, 2003; Wiley & Harnischfeger, 1974). Given the assorted definitions of time used in the prior literature the lack of consistent findings is perhaps to be expected.

Karweit and Slavin (1981) tested this possibility by examining whether time (in any form) was a significant predictor of student achievement. Eighteen elementary classrooms were recruited for this observational study, and a small subset of students (6 per class) were selected to participate. The students were observed during mathematics instruction. Four measures of time were gathered: (1) Total scheduled time (i.e. amount of time allocated to math instruction), (2) Total instructional time (i.e. amount of time spent on math instruction subtracting time for classroom procedures), (3) Engaged time (i.e. amount of time students spent engaged with the instructional activity, or time on-task), and (4) Rate of engagement (ratio of engaged time to total instructional time). Karweit and Slavin found that only student-level measures of time (i.e. engaged time and rate of engagement) were significant predictors of achievement for second and third-graders while neither student-level nor classroom-level measures of time were significant predictors of achievement for fourth and fifth-graders. The results were taken to suggest that student-level measures of time, which may more accurately reflect how students utilise instructional time, are better predictors of achievement than general classroom-level measures of time (i.e. time allocated for instruction or total instructional time).

Deploying a consistent student-level measure of time may serve to reduce variability in the strength of the relationship between time and learning; however, additional research is needed as the small sub-set of students observed per classroom may have obscured the contribution of classroom-level factors. Furthermore, lack of follow-up studies addressing this question makes it difficult to establish whether results can be generalised to present-day elementary school students.

Increasing rates of adoption of learning technologies make large amounts of data about student learning readily accessible to researchers (Kovanovic et al., 2015). Accessibility of measures of time on-task via learning technologies has renewed interest in the question of whether time is a good predictor of learning. Here we utilise an existing data set (Godwin et al., 2016) to examine whether a positive relationship between on-task behaviour and learning can be *consistently* obtained (across multiple classrooms) when utilising a uniform student-level measure of time, fraction of on-task behaviour, as student-level measures are hypothesised to yield a stronger relationship between time and learning (Karweit & Slavin, 1981). Thus, in line with the prior literature we expect to obtain evidence of a significant, positive, and non-trivial relationship between time and learning in the present study.

Various student and school level factors have been hypothesised, or found, to be associated with students' patterns of attention allocation. Here we discuss three factors identified in the prior literature: student gender, grade-level, and school type. First, researchers have observed differential patterns of attention allocation based on student gender. On average, female elementary school students tend to exhibit more on-task behaviour than male students (e.g. Godwin et al., 2016; Marks, 2000). However, it remains an open question whether the relationship between learning and the fraction of on-task behaviour is moderated by student gender. Second, patterns of attention allocation may also differ across grade-levels given that the ability to maintain a state of selective sustained attention improves with age (e.g. Bartgis et al., 2008; see Fisher & Kloos, 2016 and White, 1970 for review)—although note that Godwin et al. (2016) found an inconsistent effect of grade on students' rates of on-task behaviour across two observational studies. Therefore, more research is needed to elucidate if and how grade-level is related to students' patterns of attention allocation in classroom settings. The complex relationship between grade-level, patterns of attention allocation, and achievement is further highlighted by the findings of Karweit and Slavin (1981), reviewed above, in which engaged time and rate of engagement were significant predictors of achievement, but only for younger students. Finally, there is also some evidence that rates of on-task behaviour may differ across school types. For example, higher rates of on-task behaviour were observed in parochial elementary schools compared to private elementary schools (Godwin et al., 2016, Study 2) perhaps reflecting differences in school norms and culture. Based on these findings, we control for student gender, grade-level, and school type in our models, and also evaluate whether the relationship between learning and fraction of on-task behaviour is moderated by these factors.

Method

This data set was part of a larger study (Godwin et al., 2016) examining patterns of attention allocation in elementary students including temporal patterns across the school year and the role of student characteristics and instructional design on students' attention allocation; these findings (which do not include findings related to learning outcomes presented here) are reported elsewhere (Godwin et al., 2016).

Participants

The data set contained observational data from twenty classrooms. All students present in the classrooms were observed. One classroom was excluded from analyses due to constraints of the learning outcomes provided (see below). Classrooms were from seven schools (four public charter, three private schools) in or near a medium-sized city in the Northeastern United States. Elementary students were specifically targeted due to the protracted developmental trajectory of selective sustained attention (Bartgis et al., 2008; Fisher & Kloos, 2016), which may make this population more susceptible to engaging in off-task behaviour. At each participating school, all K-4 classrooms were invited to participate. Four grade-levels participated: five kindergarten,

five first-grade, seven second-grade, and three fourth-grade classrooms. Third-grade teachers did not volunteer to participate. The precise reason(s) why third-grade teachers were less likely to participate are unknown, one possibility is that third-grade teachers are less likely to volunteer for research due to added pressure they may feel to prepare students for the onset of standardised testing.

The final sample consisted of 19 classrooms and 356 students. Similar to prior studies in the literature, this study recorded gender as a binary variable. There were 167 male and 189 female students in the study. The mean number of students observed in a single observation session was 18.7 ($SD = 3.3$). The number of students observed per session ranged from 10 to 23 students.

Design and procedure

The data set contains two observations per classroom providing more stable and reliable estimates of the degree of students' on-task behaviour by obtaining more data on each student. The average delay between sessions was 3.05 calendar days (Range: 1–7 days). Each session lasted approximately one hour. Similar to Karweit and Slavin (1981), all observations were scheduled during mathematics instruction. Obtaining data during instruction of a single content area minimises possible interactions between student engagement and the content area. Instructional activities were determined by the teachers; thus, coders observed genuine instructional practices. The mean number of observations per session was 249.6 ($SD = 29.1$) and the mean number of times a student was observed within a session was 15.5 ($SD = 6.4$). Observations occurred between October and December 2012.

Coding behaviour

This study used the *Baker-Rodrigo Observation Method Protocol* (BROMP; Ocumpaugh et al., 2015). BROMP is a tool for conducting quantitative observations in field settings. Prior to the start of the study, coders were trained in using BROMP by coding a mixture of videotapes and live classroom observations. Cohen's Kappa values (calculated to estimate coder agreement) ranged from 0.79 to 0.84. Cohen's Kappa of 0.75 was described by Fleiss (1981) as 'excellent' in field settings; thus, coder agreement in this study exceeded the level of agreement considered excellent in field settings.

Observers used a round-robin coding strategy; this approach helps to reduce bias as it prevents observers from attending solely to more conspicuous forms of off-task behaviour. The observation order of the students was established at the beginning of every session. Each student was observed individually for up-to 20 s or until the first unequivocal behaviour was observed. The student's behaviour was then recorded on a hand held Android computer using the HART App. The observer would then proceed to code a new student following the pre-determined student observation order. The round-robin observations continued for the entire duration of the session. As a result, each student was observed multiple times over the course of the session. Importantly, this approach has been utilised in the prior literature to reliably code student affect and behaviour (Ocumpaugh et al., 2015).

First, coders used students' eye gaze to classify the students' behaviour as on- or off-task. Students were coded as being on-task if a student was looking at the teacher or instructional materials; otherwise the students were coded as being off-task. Eye gaze is a widely used measure of attention (e.g. for review see Henderson & Ferreira, 2004; Just & Carpenter, 1976). Although, we acknowledge that looking does not *always* indicate attending (e.g. mind wandering or looking elsewhere while listening to the teacher); we contend that utilising eye gaze as a proxy for attention during formal math instruction is an appropriate choice given that math instruction often contains visual elements and instructional aids that must be attended to visually. Because coders were present in the classroom, they also used contextual cues (e.g. student comments or teacher instructions) to help determine if a student was on- or off-task.

Next, if a student was coded as being off-task, the researchers classified the type of off-task behaviour using one of four mutually exclusive categories: Peer (off-task peer interactions), Environmental (off task behaviours directed at elements of the classroom environment), Self (off-task behaviours aimed at students' own clothing or other personal effects), and Other (off-task behaviours that did not unambiguously fall into the other pre-established categories, such as walking or sleeping). Findings pertaining to the prevalence of different types of off-task behaviour were reported in Godwin et al. (2016); here we utilise a subset of the data (i.e. only those classrooms that provided learning data) to focus on the relationship between overall time on-task and learning outcomes. For each student, the fraction of on-task behaviour was calculated by taking the number of times each student was on-task divided by the total number of observed behaviours (on-task and off-task).

Learning measures

Authentic learning measures were collected including: quizzes, report cards, and/or fall and winter NWEA's *Measures of Academic Progress*[®] (MAP[®]), which are used to measure academic growth of students via a computer-based assessment (Wise et al., 2013). Although utilising authentic learning measures that teachers deploy as part of their typical instructional practice helps to increase ecological validity, it also introduces variability across classrooms, a limitation we revisit in the Discussion.

The following learning outcomes were available for participants in this study: 203 students had learning outcomes consisting of quiz scores and fall and winter MAP[®] scores; four students had learning outcomes consisting of fall and/or winter MAP[®] scores; 132 students had learning outcomes consisting of report cards and quiz scores; and 17 students had learning outcomes consisting of report cards. We chose to exclude quiz scores from the analyses reported below because they were largely non-informative. Specifically, most quiz scores had low variability, with 79% of quiz scores indicating accuracy at or above 80%. Some quizzes were also based on too few items to provide a meaningful measure of student learning. It is important to note that excluding quiz scores did not affect the overall pattern of findings (see Results section for additional details). Additionally, one classroom utilised letter grades for their report cards (as opposed to percentages), leaving too few values to compute the same type of analysis as the other classrooms. Consequently, this classroom was dropped from

the analysis. As a result, both classroom observations and learning outcomes were available for 356 students from 19 classrooms.

Scores from the remaining learning measures (report cards, *MAP*[®] scores) were converted into Z-scores (for each classroom we created Z-scores for report cards and Z-scores for *MAP*[®] scores). For students with multiple learning measures, Z-scores were averaged together to create the composite variable *Learning Score* for each student. If students only had a single score, the Z-score for that measure (report card or *MAP*[®] score) was utilised as their *Learning Score*. Z-scores were computed separately for each classroom (we subtracted the classroom mean from each student and divided by the standard deviation for the classroom) to correct for different grading practices across schools and classrooms. Learning Z-scores ranged from -2.82 to 2.03 .

Analytic approach

A hierarchical linear model was run in R (version 3.4.3) (utilising lmer and lmerTest) to assess whether students with a greater fraction of on-task behaviour tended to perform better on the learning outcome measure relative to their classmates. HLM was selected as the analytic approach as it takes into account the nested or hierarchical structure of the data (i.e. students nested in classrooms). Using data aggregated at the classroom level (i.e. non-hierarchical data), correlations were calculated to assess the strength and stability of the association between fraction of on-task behaviour and learning outcomes across classrooms. We also assessed whether the relationship between learning and fraction of on-task behaviour was moderated by gender, grade level, and school type. Results are reported below.

Results

On-task behaviour

In this study, the proportion of on-task behaviour was fairly high ($M = .75$, $SD = .13$), but still consistent with estimates (.50 – .75 range; Karweit & Slavin, 1981) and observations from the prior literature (Fisher et al., 2014; Godwin et al., 2016; Godwin & Fisher, 2014; Karweit & Slavin, 1981; Lee et al., 1999; Lloyd & Loper, 1986). There was also considerable individual variability in the rates of on-task behaviour, with some students showing low rates of on-task behaviour while other students were consistently on-task (proportion of on-task behaviour ranged from .27 to 1.00).

Effect of on-task behaviour on learning

A hierarchical linear model was run in R (version 3.4.3 utilising lmer and lmerTest) to assess whether students with a greater fraction of on-task behaviour tended to perform better on the learning outcome measure relative to their classmates. *Learning Score* was entered as the dependent variable. The model included a random classroom intercept and four predictors: fraction of on-task behaviour, student gender, school type (private, public charter), and grade-level (Kindergarten, First-grade, Second-grade, Fourth-grade).

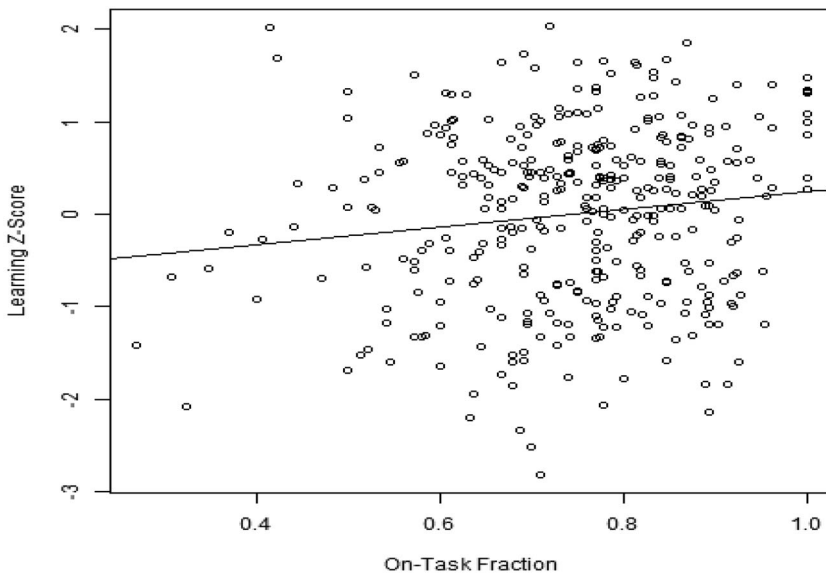


Figure 1. Association between students' fraction of on-task behaviour and learning. *Note.* Scatter plot depicting the positive relationship between the fraction of on-task behaviour and the composite variable *Learning Score* (Z-score).

On-task behaviour was a significant predictor of *Learning Scores*, controlling for grade-level, gender, and school type ($b = 1.00$, $t(349) = 2.61$, $p = .01$; see Figure 1). At the same time, this relationship was weak, with *Learning Score* increasing only by 0.20 *SD* for every 20% rise in on-task behaviour. Furthermore, a small but significant correlation was found between on-task behaviour and *Learning Score* after correcting for gender, school type, and grade-level (*partial* $r = .138$, $p = .009$), with on-task behaviour only accounting for 1.9% of the variability in *Learning Score* (R^2 difference: $0.01947 - 0.0004 = 0.019$). It is important to note that parallel findings were obtained when quiz scores were included in the *Learning Score* composite: On-task behaviour continued to be a significant predictor of student learning, controlling for gender, grade-level, and school type ($b = 0.748$, $t(368) = 2.16$, $p = .03$).

As mentioned above, the learning measures are variable across classrooms and schools; only a small subset of schools provided standardised measures (*MAP*® scores). Gathering standardised learning measures should be a focus of future research. Nevertheless, a similar pattern was obtained in a post-hoc analysis with only those classrooms that provided standardised learning measures. When *MAP*® scores were utilised as the sole dependent variable, on-task behaviour remained a significant predictor of learning after controlling for gender and grade-level (Fall *MAP*: $b = 11.48$, $t(106) = 2.03$, $p = .04$; Winter *MAP*: $b = 13.87$, $t(130) = 2.5$, $p = .01$). Again a small but significant correlation was found between on-task behaviour and *MAP* scores, after correcting for gender and grade-level (Fall *partial* $r = 0.142$, $p = .043$; Winter *partial* $r = 0.18$, $p = .01$). Including on-task behaviour as a predictor in the model accounts for less than 1% of the variability in learning scores (R^2 difference Fall *MAP* ($0.8099 - 0.8060$) = 0.0039 ; R^2 difference Winter *MAP* ($0.7943 - 0.7874$) = 0.0069).

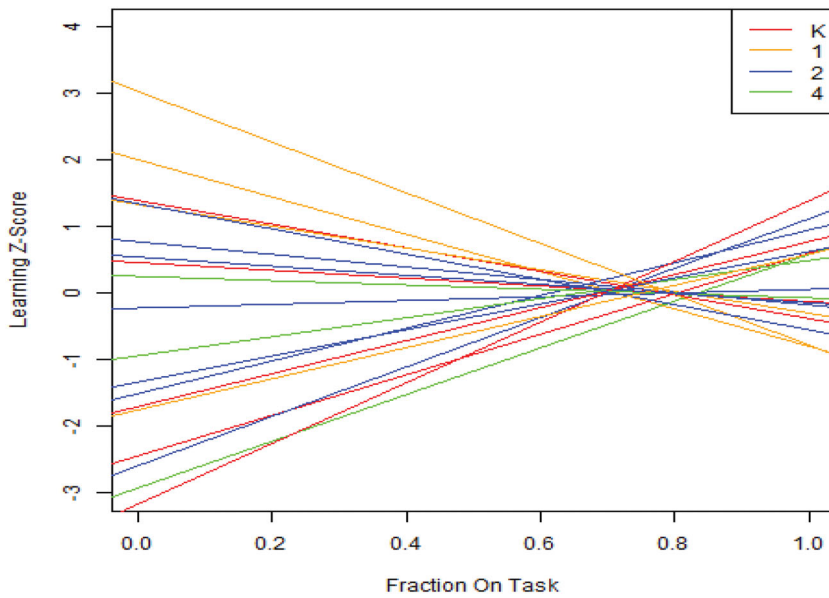


Figure 2. Classroom Regression Estimates for the relationship between the fraction of on-task behaviour and Learning. *Note.* Displays the regression estimates for the relationship between the fraction of on-task behaviour and *Learning Score* (Z-score) for each classroom (each line represents a classroom).

Effect of classroom

This study measured *time* in the same way across all participants (as the fraction of on-task behaviours); nevertheless, we found that on-task behaviour was only weakly related to learning outcomes ($r = .134$). Furthermore, we observed considerable variability in the strength of this relationship across classrooms, with slope estimates being positive for 11 classrooms and negative for eight classrooms (slope estimates for individual classrooms are displayed in Figure 2). Slope estimates reached significance in only 3 (out of 19) classrooms and in all three cases the direction of the relationship between on-task behaviour and Learning Scores was positive indicating that higher fraction of on-task behaviour was associated with better learning outcomes: one kindergarten classroom ($b = 4.54$, $p = .0026$), one Grade 2 classroom ($b = 2.46$, $p = .024$), and one Grade 4 classroom ($b = 3.51$, $p = .0238$). However, after applying Bonferroni correction for multiple comparisons, the relationship between on-task behaviour and learning outcomes remained significant only for the kindergarten classroom.

Moderation analyses

Results from prior literature indicate student and school characteristics influence rates of on-task behaviour (e.g. Godwin et al., 2016; Karweit & Slavin, 1981; Marks, 2000); thus, we further examined if the relationship between on-task behaviour and learning in this study was moderated by student grade-level, gender, and school type. There

was *no* evidence of moderation in the present data set; nevertheless the results are reported below.

Effect of gender

Prior studies reported that female students tend to exhibit more on-task behaviour than male students in elementary school (Godwin et al., 2016; Marks, 2000). In the present data set, Females exhibited numerically higher rates of on-task behaviour ($M = .76$, $SD = .13$) than males ($M = .73$, $SD = .13$); however, there was no effect of gender on *Learning Score* ($b = -.07$, $t(349) = -0.71$, $p = .48$). Additionally, the relationship between on-task behaviour and *Learning Score* was not moderated by gender ($b = .72$, $t(348) = .94$, $p = .35$). In other words, we found no evidence the relationship between learning and time on-task was stronger for females than for males (or vice versa) for any fixed school type and grade. Note, the estimate of the variance of the random intercept was zero; thus, the random effect was dropped from this model.

Effect of grade

Prior studies reported inconsistent effects of grade on on-task behaviour (Godwin et al., 2016; Karweit & Slavin, 1981). In this data set, the relationship between on-task behaviour and students' *Learning Score* was not moderated by grade-level ($F(3, 346) = 2.09$, $p = .10$). Note that the estimate of the variance of the random intercept was zero; thus, the random effect was dropped from this model. The finding that grade-level did not moderate the relationship between on-task behaviour and learning is inconsistent with the findings obtained by Karweit and Slavin (1981). In future research it will be important to sample a larger number of classrooms per grade-level from a larger sample of schools to evaluate possible grade-level effects across different samples.

Effect of school type

Similar to gender and grade, the relationship between on-task behaviour and *Learning Score* was not moderated by school type ($b = -1.51$, $t(348) = -1.36$, $p = .17$). Note, the estimate of the variance of the random intercept was zero; thus, the random effect was dropped from this model.

Collectively, the findings reported above suggest that uniformity in the measurement of time does not reduce the variability in estimates of the relationship between time on-task and learning outcomes. Therefore, these findings highlight the complex relationship between time on-task and learning, and call for more research to identify factors or conditions where more time on-task yields better learning outcomes.

Discussion

We used an existing data set to evaluate the possibility that the variability observed in the relationship between time and learning is due to differences in how time was

measured in prior research. This work extends past efforts by employing a uniform measure of time with a large sample of elementary students ($N=375$), while controlling for pertinent student characteristics (gender, grade-level) and school based factors (school type). As we expected based on the prior literature, the relationship between time on-task and learning outcomes was positive and significant; however, counter to our expectations this relationship was weak despite using a consistent and proximal measure of time (fraction of students' on-task behaviour). Furthermore, at the classroom-level both the variability in the strength of the relationship and the direction of the effect suggest divergent results obtained in the literature are not driven solely by differences in how time is measured.

Prior research posited that more proximal measures of time would yield a strong relationship between time and learning. This prediction was not borne out in the present study. Despite using a uniform student-level measure of time (fraction of on-task behaviour indexed by eye gaze), a weak relationship between time and learning was still obtained. It is possible that even this student-level measure of time does not sufficiently capture on-task behaviour, thus contributing to the weak relationship between time and learning. Future work could examine other technologies that measure attention with increased precision to assess this possibility (Godwin & Fisher, 2018).

It is also possible that time on-task is necessary but not sufficient for learning to occur or perhaps there are threshold effects in which certain amounts of on-task behaviour are needed. Future research will examine these possibilities. Results also suggest the effect of time on learning is not moderated by student gender, grade-level, or school type. Other factors may affect the strength of the relationship between time and learning. We briefly discuss these possibilities below.

The inconsistency in the strength *and* direction of the effect between on-task behaviour and learning points to a potential role for classroom-level factors not examined in this study. Teacher effectiveness may be one classroom-level factor that can strengthen or attenuate the relationship between time and learning. For example, classrooms in which students exhibit high rates of on-task behaviour and high learning outcomes may be reflective of high-quality or experienced teachers, conversely classrooms in which students exhibit high rates of on-task behaviour but minimal learning gains may indicate instruction in which teachers are entertaining but perhaps ineffectual. Teacher practices such as the quality of the delivery of instruction may be an important determinant of student learning, although prior research has noted an inconsistent relationship between teacher quality and student learning (Goe, 2007).

Differences in the strength of the observed relationship between on-task behaviour and learning may also arise from other classroom-level factors (e.g. classroom composition, classroom size, content, type of learning activity, task difficulty, pedagogy) as well as other pertinent individual differences including (but not limited to) student aptitude, prior knowledge, interest, motivation, and goal orientation. These possibilities remain to be systematically explored.

It is important to note several limitations of the present study. Due to the nature of the learning outcomes (standardised test scores, grades), these measures were not collected immediately following the observations. Therefore, it is possible students' achievement scores were also affected by additional learning opportunities occurring

at home or in subsequent lessons. Additionally, when using general measures of achievement there is not necessarily a one-to-one mapping between the instructional content presented during the observation and the learning measures. Despite this limitation, standardised measures convey important advantages; namely they obviate concerns regarding different grading practices across classrooms and schools. Future work could also include proximal learning measures to assess if a more stable relationship between on-task behaviour and learning can be obtained when using measures more closely yoked to observations of student behaviour (e.g. class assignments, unit test scores, homework grades) compared to more general measures of achievement (e.g. standardised tests).

A recent laboratory study was able to address the issue above by collecting learning outcomes data immediately after conducting observations of student on-task behaviour (Godwin & Fisher, 2014). Even when using an immediate measure of learning, the relationship between time off-task and learning was weak (adjusted $R^2 = 10\%$). However, incorporating pertinent individual difference factors (verbal IQ, processing speed, working memory, inhibitory control) in addition to time off-task yielded a model that accounted for more variability in learning (adjusted $R^2 = 39\%$). Interestingly, once individual difference factors were incorporated into the model, time off-task was no longer a significant predictor—again highlighting the need for more systematic research to examine the role of individual differences. Executive functions may be a promising starting point given the aforementioned findings and prior work documenting that executive functions are related to academic achievement (Best et al., 2011; Cragg et al., 2017; McClelland et al., 2007; Moffett & Morrison, 2020; Schmitt et al., 2017). Indeed, Moffett and Morrison (2020) found that children with weaker working memory, a component of executive function, were more likely to engage in specific types of off-task behaviour, namely passive disengagement, and proportion of time spent in passive disengagement predicted reduced reading comprehension gains in first grade.

Lastly, we did not dictate the instructional activities teachers employed, but rather elected to observe genuine instructional activities that were part of teachers' standard practice. This approach increases ecological validity but introduces variability in the types of learning activities observed. An important area for future research will be to investigate the nature of the learning activities students complete as the benefit of time may vary across different activities. As discussed previously, recent research on the 'doer effect' with adults suggests both passive tasks, such as reading, and completing instructional activities can improve learning, yet passive tasks were less efficient (Carvalho et al., 2017). Future research should assess whether variability in the relationship between time and learning observed in elementary classrooms can be accounted for in part by the types or quality of instructional activities deployed.

Summary

Despite using a uniform student-level measure of time (i.e. proportion of on-task behaviour), a weak and highly variable relationship between time and learning was observed. Previously, it was suggested researchers should '*move beyond the now well-established relation between time on task ... and student learning*', noting no further

research into this issue is warranted (Brophey, 1979, p. 743). Contrary to this claim, this study shows the relationship between time and learning is still poorly understood, over 30 years later. The current findings highlight the need for researchers and educators to re-examine this multifaceted relationship.

This is an important area of inquiry with implications for policy. The present findings should give educators and policy makers pause as they consider increasing instructional time by lengthening the school day or cutting recess in hopes of increasing students' learning outcomes. If time alone is a poor predictor of learning, simply adding more instructional time is unlikely to achieve the desired results. Nevertheless, the belief that improving academic achievement is as simple as increasing instructional time is pervasive and evident in current trends in how schools are allocating instructional time. According to a large survey conducted by the Centre on Education Policy (McMurrer, 2007), 44% of elementary schools surveyed report decreasing time allocated for physical education, art, music, and recess (among other subjects) in order to increase instructional time for core subject areas (i.e. mathematics, language arts) (for related discussion see Jarrett, 2013; Trost & van der Mars, 2010; Wilkins et al., 2003). However, these findings suggest increasing instructional time in and of itself may not yield meaningful gains in achievement given the weak and unstable relationship between time and learning. Future research is needed to identify student, classroom, and school-based factors that interact with time and learning. It is critically important to understand the conditions by which more time will aid learning, as this work will have direct instructional implications.

In sum, the relationship between time and learning was not found to be stable across classrooms, which suggests the mixed results reported in the prior literature cannot simply be attributed to issues of measurement. This work underscores the need for future research to identify circumstances in which increasing time on-task yields tangible benefits for learning.

Acknowledgments

We thank Megan Miller, Laura Pacilio, and Jessica Meeks for their help collecting data. We thank Emma Gurchiek for her help with the literature table. We also thank the children, parents, and teachers who made this project possible.

Funding

This work was supported by a Graduate Training Grant awarded to Carnegie Mellon University by the Department of Education [R305B090023] and by the Institute of Education Sciences, U.S. Department of Education [R305A110444]. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

ORCID

Karrie E. Godwin  <http://orcid.org/0000-0003-0127-986X>

Howard Seltman  <http://orcid.org/0000-0002-7543-3578>

Ryan S. Baker  <http://orcid.org/0000-0002-3051-3232>

Anna V. Fisher  <http://orcid.org/0000-0002-6661-7922>

References

- Arlin, M., & Roth, G. (1978). Pupils' use of time while reading comics and books. *American Educational Research Journal*, 15(2), 201–216. <https://doi.org/10.2307/1162460>
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004, April 24–29). *Off-task behavior in the cognitive tutor classroom: When students "game the system"* [Paper presentation]. Conference on Human Factors in Computing Systems, Vienna, Austria. <http://www.columbia.edu/~rsb2162/p383-baker-rev.pdf> <https://doi.org/10.1145/985692.985741>
- Baker, D. P., Fabrega, R., Galindo, C., & Mishook, J. (2004). Instructional time and national achievement: Cross-National evidence. *PROSPECTS*, 34(3), 311–334. <https://doi.org/10.1007/s11125-004-5310-1>
- Bartgis, J., Thomas, D. G., Lefler, E. K., & Hartung, C. M. (2008). The development of attention and response inhibition in early childhood. *Infant and Child Development*, 17(5), 491–502. <https://doi.org/10.1002/icd.563>
- Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, 21(4), 327–336. <https://doi.org/10.1016/j.lindif.2011.01.007>
- Brophey, J. E. (1979). Teacher behavior and its effects. *Journal of Educational Psychology*, 71(6), 733–750. <https://doi.org/10.1037/0022-0663.71.6.733>
- Caldwell, J. H., Huitt, W. G., & Graeber, A. O. (1982). Time spent in learning: Implications from research. *The Elementary School Journal*, 82(5), 471–480. <https://doi.org/10.1086/461282>
- Carroll, J. (1963). A model of school learning. *Teachers College Record*, 64, 723–733.
- Carvalho, P. F., McLaughlin, E. A., & Koedinger, K. R. (2017). Is there an explicit learning bias? Students beliefs, behaviors and learning outcomes. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society* (pp. 204–209). Cognitive Science Society.
- Choudhury, N., & Gorman, K. S. (2000). The relationship between sustained attention and cognitive performance in 17–24 month old toddlers. *Infant and Child Development*, 9(3), 127–146. [https://doi.org/10.1002/1522-7219\(200009\)9:3<127::AID-ICD225>3.0.CO;2-5](https://doi.org/10.1002/1522-7219(200009)9:3<127::AID-ICD225>3.0.CO;2-5)
- Cobb, J. A. (1972). Relationship of discrete classroom behavior to fourth-grade academic achievement. *Journal of Educational Psychology*, 63(1), 74–80. <https://doi.org/10.1037/h0032247>
- Connor, C. M., Jakobsons, L. J., Crowe, E. C., & Meadows, J. G. (2009). Instruction, student engagement, and reading skill growth in Reading First classrooms. *The Elementary School Journal*, 109(3), 221–250. <https://doi.org/10.1086/592305>
- Cooper, H., Robinson, J. C., & Patall, E. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research*, 76(1), 1–62. <https://doi.org/10.3102/003465430776001001>
- Cooper, H., Allen, A. B., Patall, E. A., & Dent, A. L. (2010). Effects of full-day kindergarten on academic achievement and social development. *Review of Educational Research*, 80(1), 34–70. <https://doi.org/10.3102/0034654309359185>
- Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, 162, 12–26. <https://doi.org/10.1016/j.cognition.2017.01.014>
- Edminston, R. W., & Rhoades, B. J. (1959). Predicting achievement. *The Journal of Educational Research*, 52(5), 177–180. <https://doi.org/10.1080/00220671.1959.10882563>
- Everston, C. M., Emmer, E. T., & Clements, B. S. (1980). *The junior high classroom organization study: Summary of training procedures and methodology (R&D Rep. No. 6101)*. Research and Development Center for Teacher Education.
- Fisher, A. V., Godwin, K. E., & Seltman, H. (2014). Visual environment, attention allocation, and learning in young children: When too much of a good thing may be bad. *Psychological Science*, 25(7), 1362–1370. <https://doi.org/10.1177/0956797614533801>
- Fisher, A. V., & Kloos, H. (2016). Development of selective sustained attention: The Role of executive functions. In L. Freund, P. McCardle, & J. Griffin (Eds.), *Executive function in preschool age*

- children: *Integrating measurement, neurodevelopment and translational research* (pp. 215–237). American Psychological Association. <https://doi.org/10.1037/14797-010>
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). John Wiley.
- Flunger, B., Trautwein, W., Nagengast, B., Lüdtke, O., Niggli, A., & Schnyder, I. (2015). The janus-faced nature of time spent on homework: Using latent profile analyses to predict academic achievement over a school year. *Learning and Instruction*, 39, 97–106. <https://doi.org/10.1016/j.learninstruc.2015.05.008>
- Fredrick, W. C., & Walberg, H. J. (1980). Learning as a function of time. *Journal of Educational Research*, 73(4), 183–194. <https://doi.org/10.1080/00220671.1980.10885233>
- Fredrick, W. C., Walberg, H. J., & Rasher, S. P. (1979). Time, teacher comments, and achievement in urban high schools. *Journal of Educational Research*, 73(2), 63–65. <https://doi.org/10.1080/00220671.1979.10885209>
- Gobel, P. (2008). Student Off-task behavior and motivation in the CALL classroom. *International Journal of Pedagogies and Learning*, 4(4), 4–18. <https://doi.org/10.5172/ijpl.4.4.4>
- Godwin, K. E. (2015). *The relationship between selective sustained attention and learning in young children* [Unpublished doctoral dissertation]. Carnegie Mellon University.
- Godwin, K. E., & Fisher, A. V. (2018). Wigggleometer: Measuring selective sustained attention in children. In T.T Rogers, M. Rau, X. Zhu, & C.W. Kalish (Eds.), *Proceedings of the 40th Annual Meeting of the Cognitive Science Society* (pp. 410–415). Cognitive Science Society.
- Godwin, K. E., & Fisher, A. V. (2014). Selective sustained attention, the visual environment, and learning in kindergarten-age children: Preliminary results of an individual difference study. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 541–546). Cognitive Science Society.
- Godwin, K. E., Almeda, M. V., Seltman, H., Kai, S., Skerbetz, M. D., Baker, R. S., & Fisher, A. V. (2016). Off-task behavior in elementary school children. *Learning and Instruction*, 44, 128–143. <https://doi.org/10.1016/j.learninstruc.2016.04.003>
- Goe, L. (2007). The link between teacher quality and student outcomes: A research synthesis. <https://files.eric.ed.gov/fulltext/ED521219.pdf>
- Goodman, L. (1990). *Time and learning in the special education classroom*. SUNY Press.
- Henderson, J., & Ferreira, F. (2004). *In the interface of language, vision, and action: Eye movements and the visual world*. Psychology Press.
- Hough, D., & Bryde, S. (1996, April 8–12). The effects of full-day kindergarten on student achievement and affect [Paper presentation]. American Educational Research Association. <https://files.eric.ed.gov/fulltext/ED395691.pdf>
- Hyman, H. H., Wright, C. R., & Reed, J. S. (1975). *The enduring effects of education*. University of Chicago Press.
- Jarrett, O. S. (2013). *A research-based case for recess*. Value of Play. www.usplaycoalition.clemson.edu
- Just, M., & Carpenter, P. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8(4), 441–480. <https://doi.org/10.3758/BF03201761>
- Karweit, N. (1973). *Rainy days and Mondays: An analysis of factors related to absence from school* (Report No. 162). Center for Social Organization of Schools.
- Karweit, N. (1984). Time-On-Task reconsidered: Synthesis of research on time and learning. *Educational Leadership*, 41(8), 33–35.
- Karweit, N., & Slavin, R. E. (1981). Measurement and modeling choices in studies of time and learning. *American Educational Research Journal*, 18(2), 157–171. <https://doi.org/10.2307/1162379>
- Karweit, N., & Slavin, R. E. (1980). *Time-On-Task: Issues of timing, sampling, and definition*. Center for Social Organization of Schools.
- Karweit, N., & Slavin, R. E. (1982). Time-On-Task: Issues of timing, sampling, and definition. *Journal of Educational Psychology*, 74(6), 844–851. <https://doi.org/10.1037/0022-0663.74.6.844>
- Kovanovic, V., Gasevic, D., Dawson, S., Joksimovic, S., Baker, R. S., & Hatala, M. (2015, March 16). Penetrating the Black Box of Time-on-task Estimation [Paper presentation]. International

- Learning Analytics and Knowledge Conference, Poughkeepsie, NY. <https://doi.org/10.1145/2723576.2723623>
- Kupiainen, S., Vainikainen, M., Marjanen, J., & Hautamäki, J. (2014). The role of time on task in computer-based low-stakes assessment cross-curricular skills. *Journal of Educational Psychology*, 106(3), 627–638. <https://doi.org/10.1037/a0035507>
- Lahaderne, H. M. (1968). Attitudinal and intellectual correlates of attention: A study of four sixth-grade classrooms. *Journal of Educational Psychology*, 59(5), 320–324. <https://doi.org/10.1037/h0026223>
- Lee, S. W., Kelly, K. E., & Nyre, J. E. (1999). Preliminary report on the relation of students' on-task behavior with completion of school work. *Psychological Reports*, 84(1), 267–272. <https://doi.org/10.2466/pr0.1999.84.1.267>
- Lloyd, J. W., & Loper, A. B. (1986). Measurement and evaluation of task-related learning behaviors: Attention to task and metacognition. *School Psychology Review*, 15(3), 336–345. <https://doi.org/10.1080/02796015.1986.12085238>
- Marks, H. (2000). Student engagement in instructional activity: Patterns in elementary, middle, and high school years. *American Educational Research Journal*, 37(1), 153–184. <https://doi.org/10.3102/00028312037001153>
- McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology*, 43(4), 947–959. <https://doi.org/10.1037/0012-1649.43.4.947>
- McKinney, J. D., Mason, J., Perkerson, K., & Clifford, M. (1975). Relationship between classroom behavior and academic achievement. *Journal of Educational Psychology*, 67(2), 198–203. <https://doi.org/10.1037/h0077012>
- McMurrer, J. (2007). *NCLB Year 5: Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Retrieved from the Center on Education Policy [CEP] website: <https://www.cep-dc.org/displayDocument.cfm?DocumentID=312>
- Moffett, L., & Morrison, F. J. (2020). Off-task behavior in kindergarten: Relations to executive function and academic achievement. *Journal of Educational Psychology*, 112(5), 938–955. <https://doi.org/10.1037/edu0000397>
- Odell, C. W. (1923). The effect of attendance upon school achievement. *The Journal of Educational Research*, 8(5), 422–432. <https://doi.org/10.1080/00220671.1923.10879420>
- Ocupaugh, J., Baker, R. S., & Rodrigo, M. M. T. (2015). *Baker Rodrigo Ocupaugh Method Protocol (BROMP) 2.0. Technical and Training Manual*. Teachers College, Columbia University and Ateneo Laboratory for the Learning Sciences. https://d1wqtxts1xzle7.cloudfront.net/36773439/BROMP_2.0_Final.pdf?1424899457=&response-content-disposition=inline%3B+filename%3DBaker_Rodrigo_Ocupaugh_Monitoring_Proto.pdf&Expires=1605034809&Signature=RmTm52L2b5RtY6FL24F~icT6yo3ZX54Z8bJbpUP6s1~oFrETtJr6pe4YXpullzH89MFO4FEX2rz58BSrtGC0gP5WIEf~zsqnhFQbM9kZVUm2YpQ3YhnPKWfg7AFDbhYQFb6rFOOWhRtv-LMvklgbjG~uF-NJDT5xiMhyDA18fDWAGa5Z-mz8AhGktqqCVW4Q-0lieQHvIGVc7eHVghanAeUhV~KJGYth-tli~nA8venp0vCqQ~dKI2Bdx3F5n1Fc08I-KJnyTQibSry9dQlZ6BgR0NVhbweKWdOY0cCl~gHkg3ta3grde7Md3T7lo3iTXYTI6xjXr-IL3XkAls6VQ__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
- Roberge, D., Rojas, A. J., & Baker, R. S. (2012, April 29). *Does the length of time off-task matter?* [Paper presentation]. International Conference on Learning Analytics and Knowledge, Vancouver, BC. <https://doi.org/10.1145/2330601.2330657>
- Roby, D. E. (2003). Research on school attendance and student achievement: A study of Ohio schools. *Educational Research Quarterly*, 28(1), 4–15.
- Rozelle, R. M. (1968). The relationship between absenteeism and grades. *Educational and Psychological Measurement*, 28(4), 1151–1158. <https://doi.org/10.1177/001316446802800415>
- Samuels, S. J., & Turnure, J. E. (1974). Attention and reading achievement in first-grade boys and girls. *Journal of Educational Psychology*, 66(1), 29–32. <https://doi.org/10.1037/h0035812>
- Schmitt, S. A., Geldhof, G. J., Purpura, D. J., Duncan, R., & McClelland, M. M. (2017). Examining the relations between executive function, math, and literacy during the transition to

- kindergarten: A multi-analytic approach. *Journal of Educational Psychology*, 109(8), 1120–1140. <https://doi.org/10.1037/edu0000193>
- Smith, N. M. (1979). Allocation of time and achievement in elementary social studies. *The Journal of Educational Research*, 72(4), 231–236. <https://doi.org/10.1080/00220671.1979.10885161>
- Trost, S. G., & van der Mars, H. (2010). Why we should not cut P.E. *Health and Learning*, 67(4), 60–65.
- White, S. (1970). Some general outlines of the matrix of developmental changes between 5 and 7 years. *Bulletin of the Orton Society*, 20(1), 41–57. <https://doi.org/10.1007/BF02653636>
- Wiley, D. E., & Harnischfeger, A. (1974). Explosion of a myth: Quantity of schooling and exposure to instruction, major educational vehicles. *Educational Researcher*, 3(4), 7–12. <https://doi.org/10.3102/0013189X003004007>
- Wilkins, J. L. M., Graham, G., Parker, S., Westfall, S., Fraser, R. G., & Tembo, M. (2003). Time in the arts and physical education and school achievement. *Journal of Curriculum Studies*, 35(6), 721–734. <https://doi.org/10.1080/0022027032000035113>
- Wise, S. L., Ma, L., Cronin, J., Theaker, R. A. (2013). Student test-taking effort and the assessment of student growth in evaluating teacher effectiveness. <https://www.nwea.org/content/uploads/2013/05/Student-test-taking-effort-and-teacher-effectiveness.pdf>