Classroom activities and off-task behavior in elementary school children

Karrie E. Godwin (kegodwin@andrew.cmu.edu)
Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Ma. V. Almeda (mqa2000@tc.columbia.edu)
Teachers College Columbia University, Department of Biobehavioral Sciences, 525 W. 120th St., Box 118 New York, NY 10027 USA

Megan Petroccia (petrocm@andrew.cmu.edu)
Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Ryan S. Baker (baker2@exchange.tc.columbia.edu)
Teachers College Columbia University, Department of Human Development, 525 W. 120th St., Box 118 New York, NY 10027 USA

Anna V. Fisher (fisher49@andrew.cmu.edu)
Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue, Pittsburgh, PA 15213 USA

Abstract
Maintaining focused attention in the classroom is considered an important factor for successful learning. Loss of instructional time due to off-task behavior is recognized as a significant challenge by both researchers and practitioners. However, there has been little research into the factors contributing to off-task behavior. This paper reports results from the first large-scale study investigating how elementary school children allocate their attention in classroom environments and how patterns of attention allocation change as a function of gender, grade level, and instructional format. The findings indicate that instructional format is related to off-task behavior in elementary school students. These findings can begin to form a foundation for development of research-based guidelines for instructional design aimed to optimize focused attention in classroom settings.

Keywords: Off-Task Behavior; Attention

Introduction
Loss of instructional time due to off-task behavior is a well-established problem in educational settings, recognized both by researchers (e.g., Baker, 2007; Karweit & Slavin, 1981; Lee et al., 1999) and practitioners (e.g., Lemov, 2010) for over a hundred years (cf. Currie, 1884 as cited in Berliner, 1990). The link between the quality of attention and performance has been demonstrated in the cognitive psychology literature (e.g., Choudhury & Gorman, 2006; Dixon & Salley, 2007; DeMarie-Dreblow & Miller, 1988). It has also been documented that off-task behavior has a negative impact on performance and learning outcomes in school settings (for reviews see Frederick & Walberg, 1980; Goodman, 1990).

Despite considerable prior research on off-task behavior, designing effective, easy to implement, and scalable interventions to reduce off-task behavior has been challenging. Roberts (2001) suggests that many existing interventions may be unsuccessful because they do not take into sufficient account the conditions that lead to off-task behavior. The goal of the present study was to expand upon prior research on off-task behavior in elementary school students to begin to elucidate the factors involved in off-task behavior, particularly the factors which are related to classroom activities and thus are malleable.

Off-task Behavior
There is a variety of reasons why loss of instructional time occurs in schools; these reasons include but are not limited to: weather (e.g., snow days), sudden onset interruptions (e.g., announcements over the loudspeakers), and special events. However, it has been shown that student inattentiveness (i.e., engagement in off-task behavior during instructional time) is the biggest factor that accounts for loss of instructional time (Karweit & Slavin, 1981). Prior research examining the frequency of off-task behavior has estimated that children spend between 10% and 50% of their time off-task in regular education classrooms (Lee et al., 1999; Karweit & Slavin, 1981). Classrooms employing cognitive tutors report similar results with estimates of off-task behavior constituting 15% to 25% of instructional time (e.g., Baker, Corbett, & Koedinginger, 2004; Baker, 2007).

However, there has been limited research examining the factors associated with off-task behavior. Recently, researchers have begun to explore the role of classroom design on children’s off-task behavior. Godwin and Fisher (2011) found that classroom environments that contained relatively large amounts of visual displays (e.g., charts, posters, manipulatives) elicited more off-task behavior in kindergarten children compared to visual environments that were more streamlined. These design choices were found to hinder children’s ability to attend to the content of the lesson and reduced learning outcomes. Related findings were obtained by Barrett et al. (2012). Barrett and
colleagues took a more holistic approach to design and incorporated building factors (e.g., physical space, navigation, furniture scale, etc.), environmental elements (e.g., light, sound, temperature, air quality, etc.), as well as classroom decor (e.g., color, organization, etc.). Barrett et al. found that these design choices (in combination with pupil factors) were related to students’ later academic achievement.

Instructional format (e.g., whole-class instruction, small group instruction, etc.) is another important aspect of instructional design. Yet, little is known about the relationship between instructional format and overall rates and types of off-task behavior. The goal of the present study was to examine whether type of instruction is related to incidence of off-task behavior in elementary school students.

The Present Study

This study examines whether specific instructional strategies are associated with incidence of off-task behavior in elementary school children, both in terms of the overall amount of off-task behavior, and the form which off-task behavior takes. Towards this goal we recorded patterns of attention allocation in elementary school students during a variety of instructional activities (e.g., whole-group instruction, small-group work, etc.).

Method

Participants

Twenty-two classrooms participated in the present study. Participating classrooms were selected from 5 local charter schools. Five grade-levels were recruited: Kindergarten through fourth-grade. The distribution across the five grade-levels was as follows: 5 kindergarten classrooms, 4 first-grade classrooms, 5 second-grade classrooms, 2 third-grade classrooms, and 6 fourth-grade classrooms. The average class size was 21 students (10 males, 11 females). However, due to absences the average number of children observed in a single observation session was 18.9 children. The number of children observed per session ranged from 15 to 22.

Design and Procedure

Each classroom was observed four times during the second-half of the school year, resulting in a total of 84 observation sessions. Due to time constraints in four of the 22 classrooms only three observation sessions were conducted. The observation sessions were staggered across two time periods (Time 1: February-April 2012, Time 2: May-June 2012) with two observation sessions occurring during each time period. The average delay between observation sessions within a single time period was 3.7 days (the delay ranged from 1 to 14 days). The average delay across time periods was 73.2 days. Each observation session lasted approximately one-hour.

Operationalization of on- and off-task behavior

For the present study, focused attention was defined as a “state in which attention is directed more or less exclusively to one target or task” (Ruff & Rothbart, 1996, p.110). Focused attention was operationalized through visual engagement. If children were directing their eye gaze at the teacher (or classroom assistant), the instructional activity, or toward appropriate instructional materials, the child was classified as on-task. If the child was looking elsewhere, they were classified as off-task. Eye gaze is a common measure of visual attention (for reviews see Henderson & Ferreira, 2004; Just & Carpenter, 1976), and it is arguably a reasonable (albeit imperfect) measure of focused attention.

Coding

All coders were trained in the Baker-Rodrigo Observation Method Protocol (BROMP) for coding behavioral data in field settings (Ocumpaugh, Baker, & Rodrigo, 2012) using software developed for the android handheld computer. All coders received extensive training consisting of coding videotapes and live observation sessions. Inter-rater reliability was established prior to the study proper. Kappa values ranged from 0.79 to 0.84. This level of reliability is in line with past classroom research coding off-task behavior, and exceeded the 0.75 threshold to which Fleiss (1981) refers as “excellent” in field settings.

Children were observed using a round-robin coding strategy, in order to reduce the tendency of observers to attend to more salient instances of off-task behavior. The order in which children were observed was determined at the beginning of each session. Each time a child was observed the observation lasted for up to 20 seconds. The first unambiguous behavior observed during the 20-second period was recorded. Quick glances were considered ambiguous behaviors, and coders were instructed to wait for a clear behavior to occur. If a behavior was noted before 20 seconds elapsed, the coder proceeded to the next child, and a new 20-second observation period began. Coders observed the children using peripheral vision or side-glances. Peripheral vision was utilized in order to avoid looking directly at the student being observed. This technique makes it less apparent to the child that s/he is being observed. This procedure has successfully and reliably captured students’ behavior in prior work which assessed student behavior and affect (cf. Baker et al., 2006; Baker et al., 2010; Ocumpaugh et al., 2012).

Coders classified children’s behavior as on- or off-task. If the child was looking at the teacher (or classroom assistant), the instructional activity, and/or the relevant instructional materials, they were categorized as on-task. If the child was looking elsewhere, they were categorized as off-task. Contextual clues (i.e., teacher instructions) were also taken into consideration when distinguishing between on- and off-task behavior. For example, if a child was instructed to discuss an idea with a partner, coders would classify conversing with another peer as on-task unless the coders could clearly discern that the conversation was unrelated to the instructional task.

If the child was classified as off-task, the type of off-task behavior was recorded. Six mutually exclusive categories of
off-task behavior were logged: (1) Self-distraction, (2) Peer
distraction, (3) Environmental distraction, (4) Supplies,
(5) Walking, or (6) Other. Self-distraction entailed
engagement with something on the child’s own body, such
as an article of clothing or an appendage, as well as episodes
in which the child would close their eyes. Peer distraction
was defined as interacting with or looking at another
student(s) when not directed to do so. Environmental
distractions include interacting with or looking at any object
in the classroom that was not related to the task at hand,
while Supplies consists of inappropriately using any object
that was part of the assigned task (e.g., playing with a
writing utensil). Walking was operationalized as a student
physically walking around the classroom when it was not
considered appropriate for the task. Other distractions
included student behavior that was off-task but did not
clearly align with the five aforementioned categories. A
seventh category Unknown was also included to capture rare
instances in which it was unknown whether the child was
on- or off-task, and it was impossible or inappropriate for
the observer to relocate in order to obtain a better view of
the child. Unknown was also used when students left the
classroom for various reasons (e.g., to use the restroom).

Children in each session were treated as a different set of
students since it was not possible to link observations across
the four sessions. Thus, a total of 1,587 student-session pairs
were observed. A student-session pair refers to a specific
student observed by a coder within a specific session. The
average number of observations per session was 330.13 and
the average number of observations per child within a
session was 17.58.

Data Analysis: Variables

Using these data, we attempted to predict within an
observation session each student’s total on- or off-task
behavior as well as the type of behavior the student tended
to engage in while off-task. Two categories of predictor
variables were considered for incorporation into the models:
student characteristics and instructional design. Gender and
grade were included as student characteristics. Predictor
variables pertaining to instructional design included the
proportion of each classroom instructional format and the
variable Transitions/Duration of Instructional Format.

Instructional format was included as a predictor variable
in order to examine whether certain instructional formats
elicited differential amounts of off-task behavior. Six different
instructional formats were coded: (1) individual work, (2)
small-group or partner work, (3) whole-group instruction at
desks, (4) whole-group instruction while sitting on the
carpet, (5) dancing, and (6) testing. The proportion of time
students spent in each of the aforementioned formats was
calculated. The average duration for each instructional
format is provided in Table 1.

Transitions were noted every time the teacher paused
instruction to change from one activity to another (e.g.,
transitioning from working on a math problem to listening
to a short story). In many cases, transitions coincided with a
change in instructional format (e.g., switching from whole-
group instruction to small-group instruction); however this
was not always the case as transitions could occur without a
change in instructional format (e.g., with children rotating
from one small group activity to another). Transitions were
frequently marked by the teacher asking the children to get
out new instructional materials (e.g., “Please get out your
math binders’) or requesting students to change locations
(e.g., “Please put your notebooks away and come to the
carpet”).

<table>
<thead>
<tr>
<th>Average Time Spent (sec) Per Instructional Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Work</td>
</tr>
<tr>
<td>Small Group</td>
</tr>
<tr>
<td>Whole-group Instruction at Desks</td>
</tr>
<tr>
<td>Whole-group Instruction on Carpet</td>
</tr>
<tr>
<td>Dancing</td>
</tr>
<tr>
<td>Testing</td>
</tr>
</tbody>
</table>

The primary dependent variable was the proportion of on-
task behavior of a specific student within a specific session.
Additional models were also constructed in order to predict
peer off-task behavior and environment-based off-task
behavior, as these two types of off-task behavior were
common sources of distraction for elementary school
children (See Table 2). Environment-based off-task
behavior was of particular interest as it is a malleable feature
that could theoretically be targeted when designing
interventions aimed to mitigate off-task behavior.

Data Analysis: Approach

We predicted student on-task behavior using a regression
tree algorithm (cf. Witten & Frank, 2005), which sets up a
decision tree to predict a numerical value. Binary decisions
are made based on specific variables. After several decisions
are made, a numerical prediction is given. To determine
these specific variables, regression trees find breakpoints
within data, where relationships change (mostly) at a certain
value of a variable. Regression trees can find more
complicated interactions and relationships between variables
than is typically possible with linear regression methods,
while still remaining more constrained than neural networks
or support vector machines—as such, they occupy a
moderate position in the trade-off between goodness of fit
and flexibility of fit parsimony. The specific
implementation of regression tree used in this paper is
REPTree in RapidMiner 5.2 (Mierswa et al., 2006). This
relatively rapid algorithm builds a tree using reduced error
pruning; an approach designed to produce relatively
conservative models (Witten & Frank, 2005).

Resultant models were evaluated using six-fold student
level cross-validation. In this process, students are split
randomly into six groups. For each possible combination, a
feature is developed using data from five groups of students.
before being tested on the sixth “held out” group of students. By cross validating at this level, we increase confidence that features will be accurate for new students.

Within this paper, cross-validation (Efron & Gong, 1983) is used instead of statistical significance testing for multiple reasons. First, cross-validation assesses how accurate a model is likely to be for new data, rather than assessing the likelihood that a specific data set’s results are due to chance. In assessing generalizability, cross-validation has the same goal as the use of information criteria. In fact, the k-fold cross-validation approach used here is thought to be asymptotically equivalent to the Bayesian Information Criterion (BiC) (Shao, 1993). Second, there is not an appropriate statistical significance test for the data used here for two reasons: (1) there is not a well-known statistical significance test for regression trees, and (2) student IDs are not connected across sessions. Testing statistical significance without a student term would result in a bias strongly in the direction of statistical significance; conversely, using a student-session term would result in having an order of magnitude more parameters, biasing strongly against statistical significance.

### Results

Consistent with prior research, children were largely on-task: 71% of children’s observed behaviors were on-task. As seen in Table 2, three of the most common types of off-task behavior observed were Peer distractions (45%), Self-Distractions (18%), and Environmental distractions (16%).

<table>
<thead>
<tr>
<th>Table 2. Descriptive statistics for students’ on- and off-task behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proportion of Observed Behaviors</strong></td>
</tr>
<tr>
<td>On-task Behavior</td>
</tr>
<tr>
<td>Off-task Behavior</td>
</tr>
</tbody>
</table>

| **Proportion of Off-task Behaviors**                       |
| Self-Distraction                                           | 18% |
| Peer Distraction                                            | 45% |
| Environmental Distractions                                  | 16% |
| Supply Distractions                                         | 11% |
| Walking                                                     | 3% |
| Other Distractions                                          | 8% |

| **Descriptive Statistics**                                  |
| Observations per session                                    | 330.13 (63.6) |
| Observations per session per child                          | 17.58 (3.7)   |
| Student/Session pairs observed                              | 1,587         |

Models predicting on-task behavior were fit based both on instructional design and on limited student demographics (e.g., grade-level and gender). The best overall model predicting on-task behavior was found for the regression tree when student demographics were not included. This model obtained a cross-validated correlation coefficient of \( r = 0.352 \). The cross-validated correlation coefficients for the “instructional design plus demographic” models were as follows: A regression tree which added gender achieved a cross-validated correlation of 0.322 to the frequency of student on-task behavior and a regression tree model which added grade-level achieved a cross-validated correlation of 0.329. As these “instructional design plus demographics” models achieved lower cross-validated correlation than the simpler model which only considers instructional design, we can infer that including this demographic information does not improve model fit in a generalizable fashion (as mentioned above, this is akin to achieving a better BiC: the additional fit does not compensate for the added model complexity/ flexibility of fit). As such, for determining off-task behavior it does not appear to be important whether an elementary school student is a boy or a girl, once instructional design is taken into account. Similarly, grade-level does not seem to be an important factor, once the influence of grade on instructional design is taken into account.

In this data set, regression trees achieved generally better performance than linear regression. A linear regression model based on instructional design achieved a cross-validated correlation of 0.221 to the frequency of student on-task behavior. No linear regression model (regardless of the feature set used) performed better than the corresponding regression tree model.

Within instructional design, both the format and the variable Transitions/Duration of Instructional Format were associated with a better model. Removing either of these variable types from the model resulted in worse cross-validated correlation.

The final regression tree model was rather complex, with 63 leaf nodes (final decision values) and 62 decision nodes. It can be easier to understand some of the key data relationships by considering the cross-validated and regular correlations for single-feature linear regression models. In Table 2, both cross-validated correlations and regular correlations are given. It is worth noting that cross-validated correlations should always be positive (a negative cross-validated correlation does not imply a negative relationship, but that the relationship reverses direction when applied to different parts of the data; e.g., a negative cross-validated correlation implies that the model is worse than chance). Directionality of the relationship should be inferred from the regular correlation.

As seen in Table 3, the relationship between instructional format and on-task behavior varies as a function of the type of instructional format. Individual work and whole-group instruction at desks were negatively associated with on-task behavior, while small group-work, whole-group instruction while sitting on the carpet, dancing, and testing were positively associated with on-task behavior. It is worth noting that the individual variables may have weak associations, even as reasonable prediction is achieved from a combination of variables. Note, however, that this is not simply a case of an overly-complex model predicting noise; the cross-validated correlation of the overall model is an
indication that the model works on entirely unseen data. The variable Transitions/Duration of Instructional Format was also found to be positively correlated with on-task behavior.

We also generated models to predict peer off-task behavior and environmental off-task behavior, using the same features and modeling methods. The cross-validated correlation of the REPtree model based solely on instructional design was 0.244 for peer off-task behavior and 0.161 for environmental off-task behavior. These correlations did not increase substantially if gender or grade-level were included. With these results, the peer model appears to perform somewhat better than the weak correlation achieved in the environment model, but neither model was as effective as the model predicting the overall amount of on-task behavior.

Table 3. Goodness of single-feature linear regression models at predicting on-task behavior (note that cross-validated correlations are always positive, unless the model performs worse than chance on new data).

<table>
<thead>
<tr>
<th>Feature</th>
<th>Direction of relationship</th>
<th>Cross-validated correlations</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual work</td>
<td>Negative</td>
<td>.000</td>
<td>-.018</td>
</tr>
<tr>
<td>Small-group work</td>
<td>Positive</td>
<td>.000</td>
<td>.032</td>
</tr>
<tr>
<td>Whole-group instruction at desks</td>
<td>Negative</td>
<td>.114</td>
<td>-.113</td>
</tr>
<tr>
<td>Whole-group instruction carpet</td>
<td>Positive</td>
<td>.110</td>
<td>.110</td>
</tr>
<tr>
<td>Dancing</td>
<td>Positive</td>
<td>.017</td>
<td>.043</td>
</tr>
<tr>
<td>Testing</td>
<td>Positive</td>
<td>.025</td>
<td>.051</td>
</tr>
<tr>
<td>Transitions/Duration of Inst. Format</td>
<td>Positive</td>
<td>.075</td>
<td>.075</td>
</tr>
<tr>
<td>Gender</td>
<td>Positive</td>
<td>.108</td>
<td>.109</td>
</tr>
<tr>
<td>Grade</td>
<td>Positive</td>
<td>.005</td>
<td>.039</td>
</tr>
</tbody>
</table>

The strongest individual feature correlation (using linear regression and non-cross-validated correlations) for peer off-task behavior is the amount of time spent in whole-group instruction while sitting on the carpet ($r$=0.136), followed by the amount of time spent in small-group work ($r$=0.119). Similarly, the strongest individual feature correlation for environmental off-task behavior is small-group work ($r$=-0.115). These findings suggest that instructional format does matter for determining specific off-task behaviors. However, the magnitude of correlation for the full model indicates that instructional format determines to a greater degree whether a student will go off-task, than exactly how they will go off-task. Clearly, the type of instructional format may influence students’ choices of how they will go off-task (e.g., a student may be more likely to engage in peer off-task behavior during small-group work when another child is in close proximity); however, the exact manifestation of off-task behavior may be influenced by momentary factors (i.e., the most interesting item/person in the classroom at a specific moment).

**Discussion**

The present work is the first large-scale study of off-task behavior in elementary school students to investigate the relationship between features of instructional design and incidence of off-task behavior. Specifically, we examined whether type of instructional format (e.g., individual work, small-group work, whole-group work, etc.) and the variable Transitions/Duration of Instructional Format are related to the overall rate of off-task behavior. Our findings indicate that both variables are related to children’s engagement in instructional activities. At the same time, children’s gender and grade-level (K-4) made only a marginal contribution to off-task behavior once features of the instructional design (e.g., instruction format and variable Transitions/Duration of Instructional Format) were taken into account.

The reported results also indicate that certain types of instructional format are associated with more on-task behavior than others, although further research is required to explicate this finding. There are several possible hypotheses that could be explored. One potential underlying factor is variations in teacher supervision. It is feasible that classroom management is easier for certain instructional formats (e.g., small-group work) than others (e.g., whole-group instruction at desks). Consequently, instructional formats that are easier for teachers to supervise may result in a reduction in opportunities for students to go off-task.

Secondly, student engagement in the instructional task may also vary across instructional formats. For instance, instructional activities that take place individually or at the students’ desks (i.e., whole-group instruction at desks) may be less engaging or motivating than small-group activities which tend to be more socially oriented and include more hands-on learning components. Instructional activities that are more motivational may in turn increase students’ on-task behavior. Additionally, instructional duration varies across these formats. Thus, children may be better able to maintain a state of focused attention when instruction consists of small blocks of instructional activities verses instructional activities that occur over a longer duration (cf. Ruff & Lawson, 1990; Sarid & Breznitz, 1997). Currently these hypotheses are speculative, and they require additional investigation to determine their viability.

As stated previously, off-task behavior is a significant problem in educational settings because it is thought to impede learning. Optimizing instructional design to promote on-task behavior is a desirable goal; however, there is a paucity of research linking instructional design choices to
attention allocation in classroom settings. The present findings are a first-step in providing empirical evidence to inform instructional design.

Acknowledgments

We thank Laura Pacilio for her help collecting data. We also thank the children, parents, and teachers who made this project possible. The work reported here 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.

References


