



This is to certify that the
dissertation entitled
**Sources of Error in Measures of Time
Allocation in the Classroom**

presented by

David J. Solomon

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Educational Psychology

William B. Schmitt

Major professor

Date August 1, 1983



RETURNING MATERIALS:

Place in book drop to
remove this checkout from
your record. FINES will
be charged if book is
returned after the date
stamped below.

BOOK USE ONLY

DO NOT CHECK IN

SOURCES OF ERROR IN MEASURING TIME ALLOCATION
IN ELEMENTARY CLASSROOMS

By

David J. Solomon

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Educational Psychology

1983

ABSTRACT

SOURCES OF ERROR IN MEASURING TIME ALLOCATION IN ELEMENTARY CLASSROOMS

By

David J. Solomon

This dissertation investigates error in measuring classroom time allocation within the framework of the Harnischfenger and Wiley (1976) model of school learning. First, a generalizability study was done where the facets were measures (teacher logs verses observer field notes), classes, days and students. Secondly, the reliability of a coding procedure was assessed. Thirdly, the error in teacher logs was modeled.

Both the teacher as well as an outside observer recorded the activities of each student in six classes for eight days over a three month period. Two separate individuals each coded the observer descriptions for two of the classes for four days.

The variance for each of the facets in the generalizability study as well as their interactions were computed. These were used to estimate the reliability of a number of data collection designs. Mean differences and absolute value mean differences were used to assess the consistency of the multiple codings. Error in teacher recorded pursuit length was partitioned into a fixed bias, random bias and random error based on a measurement model developed by Schmidt (1981). The error in categorizing pursuit based

on

d.

ti

u

a

n

l

s

ti

ti

|

a

h

t

d.

a

s

n

i

r

l

on teacher logs was evaluated.

The results indicated that increasing the number of days observed was the most powerful approach for improving the reliability of time allocation studies. However the use of observers as opposed to teachers as a data source also resulted in a substantial improvement. Increasing the number of students observed within a class resulted in little or no improvement. The multiple codings of the same observer notes were found to be reasonably consistent. The error in teacher estimates of pursuit length was mainly random. The pursuit records coded from observer and teacher descriptions were fairly consistent in terms of how subject matters were categorized with the exception of teacher supervision.

The results suggest that teachers can collect reliable data for measuring differences among classrooms on time allocation. If the focus is on measuring differences among students within a class, or the time categories are narrowly defined as in the Harnischfieger and Wiley model, it may not be possible to obtain acceptable levels of reliability without resorting to observers and/or extremely large research designs.

ACKNOWLEDGMENTS

I would like to thank my wife, Carolyn Solomon, for the many hours of editorial assistance she provided as well as patience beyond the call of duty prior to my orals. I would also like to thank my advisor, Dr. William Schmidt, and committee member, Dr. Robert Floden, for the help and constructive criticism they provided. In addition I would like to thank my employer, Touchstone Applied Science Associates, for the use of their data processing equipment.

TABLE OF CONTENTS

	Page
LIST OF TABLES	v
CHAPTER ONE: INTRODUCTION	1
Purpose	2
Measurement of Pupil Pursuits	3
CHAPTER TWO: REVIEW OF THE LITERATURE	7
The Harnischfeger and Wiley Model	8
The Relationship Between Time and Learning	11
Engagement Rates	17
The Reliability of Teacher Recorded Data	24
CHAPTER THREE: THE DESIGN OF THE STUDY	28
Research Questions	29
Sample	31
Data Collection Procedures	31
Estimation of Variance Components	32
Procedures with Time per Day per Student as Unit	34
Method Used to Assess Error at the Pursuit Level	36
Coding Procedures	37
The Analysis of Coded Pursuit Records	40
The Analysis of the Categorization of Pursuits	42

	Page
CHAPTER THREE: CONTINUED	
Analysis of Errors in Pursuit Length	43
Summary	45
CHAPTER FOUR: THE RESULTS	48
Estimation of Variance Components	49
Reliability of the Coding Procedures	58
The Consistency of Log and Observation Estimates	68
Individual Pursuit Level Analysis	83
Pursuit Level Analyses	85
Discrepancies in Pursuit Coding	86
Discrepancies in the Length of Pursuits	97
CHAPTER FIVE: CONCLUSIONS AND IMPLICATIONS FOR MEASURING TIME ON TASK	116
Summary of the Results	116
Conclusions	124
Coder Reliability	125
The Generalizability of Measures of Time Allocation	126
The Accuracy of Teacher Logs	127
APPENDIX A	132
APPENDIX B	143
APPENDIX C	154
BIBLIOGRAPHY	157

LIST OF TABLES

Table	Page
1. Correlation of Time and Learning	12
2. Lower Elementary (Grades 2 and 3) N=33 . . .	19
3. Upper Elementary (Grades 4 and 5) N=62 . . .	19
4. Pursuit Record Data Elements	38
5. Variance Components of Measures of Five Subjects and Transitions	50
6. Generalizability Coefficients for Measuring Time Allocation	56
7. Comparison of Time Measures from Multiple Codings	60
8. Activity Time Measure Differences	70
9. Ratio of Time Measure Differences Over Sources of Variance	72
10. Differences Among Classes in Pursuit Categorization	87
11. Categorization of Activity Type from Logs and Observations	88
12. Categorization of Activity Type from Logs and Observations by Class	89
13. Categorization of Group Type from Logs and Observations	92
14. Categorization of Group from Logs and Observations by Class	93
15. Categorization of Supervision from Logs and Observations	95
16. Categorization of Supervision from Logs and Observations	96

Table	Page
17. Sources of Error in Pursuit Length Coded from Teacher Logs by Class	99
18. Bias at Standard Pursuit Lengths by Class . .	101
19. Bias and Error Components of Measurement Error Variance by Classroom	104
20. Sources of Error in Pursuit Length Coded from Teacher Logs by Activity	106
21. Bias at Standard Pursuit Lengths by Activity	107
22. Bias and Error Components of Measurement Error Variance by Activity	110

th

te

su

in

st

ti

no

it

on

re

to

an

me

19

of

ma

an

on

inc

CHAPTER ONE

INTRODUCTION

One of the most critical roles a teacher commands is the determination of pupil activities. The decisions teachers make about how pupil time is allocated to various subject matters and in what types of settings greatly influences what students learn. With this in mind, student activities seem a logical focal point for educational research.

A growing amount of educational research is in fact now focusing on student activities especially one aspect of it called "time on task." Stallings (1980) has called time on task one of the most useful variables to emerge from research on teaching in the 1970's. Much of the research to date has shown a substantial relationship between the amount of time allocated to a subject matter and achievement in that area (Wiley, 1976; Fisher, 1976; Stallings, 1980; Schmidt, 1981). There are also many other aspects of student activities that in conjunction with subject matter affect what children learn in school. Harnischfieger and Wiley (1976) in their model of school learning focused on instructional grouping (whole group, subgroup and individual) and type of supervision (whether directly

sup

the

a n

in

in

sk

to

so

so

and

bee

rec

are

is

tic

add

in

abi

per

spe

var

pro

PUR

imp

supervised by the teacher or not). Another dimension is the extent to which a single student activity integrates a number of different subject matters. Language arts instruction for example, offers a unique potential for integrating two or more curricula. Reading and writing are skills that must be taught using some content. In order to read or write, one must read or write something. That something can be another subject matter such as science or social studies allowing a student to simultaneously learn another subject matter area. Although integration has been talked about for over fifty years (Symonds, 1930), research is needed on the extent to which student pursuits are actually integrated in classrooms and if this approach is indeed effective.

There are many other interesting questions in educational research that the study of pupil activities can address. Allington (1980) found the time students spend in school reading was positively related to their reading ability. Research on the relationship between teachers' perceptions of students' abilities and the time students spend in different pursuits could help explain why the variation in achievement between students grows as they progress through school.

PURPOSE

Given the interest and growing evidence of the importance of time allocation in the classroom, research is

needed on how best to measure this variable. Collecting data on how time is spent by students in school can be a complex and expensive process. The Language Arts Project of the Institute for Research on Teaching has developed a set of procedures for collecting time allocation data at the student level within the framework of a model of school learning developed by Harnischfeger and Wiley (1976). The purpose of this dissertation is to evaluate various aspects of this procedure with the hope of determining more accurate and efficient methods of collecting student time allocation data.

Two basic methods were used to collect descriptions of student time allocation. The first were logs recorded by the teacher throughout the school day. The second method was structured field observations.

MEASUREMENT OF PUPIL PURSUITS

The procedures developed by the Language Arts Project to collect time allocation data consisted of two stages. First a description of student activities and the times they occur were to be obtained. Secondly, these descriptions were to be coded into pupil pursuit records with the beginning and ending time as well as codes indicating the subject matter, group type and supervision of the activity. Errors can occur in both describing the activities and coding them into pupil pursuit records.

There are a number of ways descriptions of student

classroom activities can be obtained. The use of observer recorded structured field notes and teacher recorded logs are evaluated by this study. It is hypothesized that observers are more accurate than teachers in that they can focus all their efforts on data collection while the teacher's main job is teaching that at times can take all of his/her effort. A teacher is likely to be recording pupil pursuits and the time they occur after they happen especially when the classroom situation is hectic. The teacher in some cases would have to estimate starting and ending times and could forget to record activities. An observer focusing his/her whole attention on data collection is less likely to miss activities or have to guess as to their beginning and ending times. Using observer transcripts as descriptions of classroom activities generally would be more expensive than using teacher logs. The costs of using teachers or observers for obtaining student activity descriptions is roughly proportional to the number of days data is collected.

The second stage of measuring the time students spend in different pursuits is coding descriptions of student activities into pupil pursuit records. As was stated above these records contain a beginning and ending time for the pursuit and codes categorizing the pursuit in terms of subject matter, group type and teacher supervision. It is hypothesized that coding error mainly consists of improperly categorizing pupil pursuits on one or more of

the dimensions. The beginning and ending times are contained in the descriptions and it is expected that it would be rare for coders to incorrectly copy them. Also this type of error is easily caught via computer error checks since it results in gaps or overlaps in pupil pursuit records.

Coder accuracy is probably to a large extent determined by the categorization scheme and the coder's training in its use. The categorization of any phenomenon is always to some extent artificial and ambiguities usually exist. Harnischfeger and Wiley's model defines the categories for group type and type of supervision but leaves it up to individual researchers on how to categorize subject matter. It is hypothesized that the greater the number of subject matter categories and the more detailed they are, the more difficult it would be to have consistency across coders. Assuming reasonable care is used in developing a coding scheme and training coders, it is hypothesized that the coding process would introduce relatively small amounts of error compared to the error from descriptions of pupil activities.

The amount of funds available for data collection in research studies is always limited. In determining the most appropriate use of resources for data collection in time allocation studies it is necessary to balance the cost of increasing the accuracy of the data on each classroom with the cost of collecting data on as many classrooms

as possible which increases generalizability and reduces sampling error. To make prudent decisions as to the best data collection design for answering a given set of research questions, researchers using the Harnischfeger and Wiley model as a paradigm need as much information as possible as to the amount of error introduced by various approaches as well as which types of student activities are most difficult to record.

CHAPTER TWO

REVIEW OF THE LITERATURE

Harnischfeger and Wiley's model of school learning (1976) has provided the framework for the data collection procedures this paper is assessing. The first section of this chapter presents their model. The model suggests that the amount of time allotted for students to learn different subject matters under different learning settings is the single most important determiner of school achievement. The second section of this chapter reviews the literature relating time on task and achievement. As was stated in chapter one, most theorists believe it is only the allotted time in school that a student spends actively attempting to learn that promotes achievement. The third section of this chapter reviews the research on engagement rates and the relationship between allocated time and engaged time in the classroom. The major focus of this study is investigating the nature and extent of errors in teacher recorded data on student time allocation. The fourth section of this chapter will discuss the research on the accuracy of teacher recorded information on classroom activities.

THE HARNISCHFEGER AND WILEY MODEL

Harnischfeger and Wiley's model is based on two assumptions. The first is that the most important factor determining student achievement on a topic is the total amount of time the student spends actively attempting to learn the topic. The second is that there are large differences in the total allocated learning time students receive in various curricula under different learning settings. Given these two assumptions, the model focuses on student classroom activities. Harnischfeger and Wiley strongly believe it is only by shaping pupil activities that factors such as curricula and teacher actions influence school learning.

There are a number of factors that control the time students spend learning different topics. The state government usually sets the number of school days in a year while the school administration controls the length of the school day, breaks such as lunch and recess as well as to some extent the curriculum. Within these limits the classroom teacher allocates time to different subject matters. The teacher also controls the structure of the learning setting under which students receive instruction on different topics. The students of course decide the amount of effort that they put into mastering a topic although influenced by teacher's actions and classroom setting.

The crucial variable in Harnischfeger and Wiley's

model is the time students spend in what they term pursuits. A pupil pursuit is defined as the intersection of three dimensions. These dimensions are subject matter, group type broken into whole class, subgroups and individual activities and whether or not the activity is directly supervised by the teacher. The model is focused on these dimensions because Harnischfeger and Wiley believe that these are the primary dimensions along which teachers organize classroom activities. The students in a class move from activity to activity throughout the school day, usually with a transition period in between activities. It is these individual student activities that form the basic unit of pupil pursuits. This model focuses on the teacher's role of allocating scarce resources. The resources are the limited amount of time in school children have to learn different subject matters and the teachers' limited time he/she must distribute among twenty or thirty students in the class.

Harnischfeger and Wiley acknowledge that teachers perform many important roles in facilitating student learning other than allocating resources. They explain material, motivate students as well as provide them with feedback as to their performance. The model focuses on time allocation because Harnischfeger and Wiley feel that is the most powerful approach to improving student achievement.

The basic premise of this model is that pupil pursuits

mediate the effects of such things as curricula, administrative policy and teacher actions in promoting student learning. Given this assertion, two broad types of research questions are suggested by the model. The first is what are the effects of different types of policy decisions and teacher practices on pupil pursuits. For example, how much academic learning time is gained by increasing the school day by an hour? Do open versus traditional classrooms result in a large increase in transition time? The second type of research question is what kinds of pupil pursuits seem to work best for which types of students. For example, is working in a small group on an experiment an effective way of teaching the scientific method to elementary school children? Is it necessary to have the teacher directly supervise and provide immediate feedback to students learning beginning reading skills?

In answering both types of research questions suggested by the model, it is necessary to measure the time students spend in different pursuits. The major focus of this dissertation is to determine whether teacher logs can provide a reasonably accurate though economical approach to collecting classroom time allocation data as compared with the use of classroom observers.

THE RELATIONSHIP BETWEEN TIME AND LEARNING

It seems inherently obvious that exposure to a given subject matter is a precondition for mastery of the skills and content contained as part of that subject matter. There are probably many other aspects of both the learner and the instructional setting that affect the extent of what students learn in school, however exposure seems necessary. For years researchers have been struck by the tremendous differences in the time students receive in different topics (Borg, 1980). This section discusses some of the empirical research that has assessed the relationship between time and learning.

The majority of studies investigating the relationship between time and achievement have found a positive relationship though the size of this relationship has varied considerably. These differences are likely due to both sampling error in the individual studies as well as the wide range of methods, conditions, subject and content areas used.

This section will begin by discussing a review of the literature relating time and learning. This will be followed by descriptions of three large scale studies that control for background variables and measure achievement at the student level while measuring time allocations at the school level.

Fredrick and Walberg (1980) reviewed approximately fifty studies relating time and learning. They categorized

the studies into those measuring instructional time in years, days, hours, and minutes. They found a consistent though moderate relationship in all four categories of studies. The correlation ranges are shown in Table 1.

Table 1
Correlation of Time and Learning

Category	Low	High
Years of instruction	.26	.71
Days of instruction	.36	.69
Hours of instruction	.13	.59
Minutes of instruction	.15	.53

As one might expect, controlling for social class depressed the correlations in a number of studies. The authors also make the point that a number of studies found the log of the instructional time tended to be a better predictor of achievement than the actual time. This suggests increases in achievement with increases in time spent on particular topics may drop off after some point.

Wiley (1976) performed a hierarchical analysis relating quantity of schooling to achievement in math, reading, and verbal ability controlling for student background variables. The analysis was hierarchical in the sense that student achievement and background were measured at the student level while quantity of schooling was measured at the school level. Wiley used a portion of the Equality of

Educational Opportunity Report (EEOR) data consisting of 2,558 sixth grade students from 40 central city Detroit schools. The achievement measures for verbal ability consisted of sentence completion and synonym subtests from Educational Testing Services School and College Ability Test series. The math and reading ability measures were from Educational Testing Services Sequential Test of Educational Progress. The measure of quantity of schooling was the log of the triple product of the average daily attendance rate, length of school day and number of school days per year. The student background variables consisted of race, number of siblings and the number of certain types of possessions in the home.

Wiley found quantity of schooling had a large effect on achievement in all three areas. From the path coefficients he obtained (4.88, 9.76, and 11.12 with standard errors of 1.62, 2.80, and 3.00 respectively for math, reading, and verbal), Wiley concluded a 24% increase in the quantity of schooling would result in a 34%, 65%, and 34% increase in verbal, reading and math scores respectively.

Karweit (1976a, 1976b) performed a set of analyses similar to Wiley's. She used 30 schools from the suburban Detroit area. Using the same analysis procedures as Wiley, she found nonsignificant ($p \geq .05$) effects for quantity of schooling on all three measures (math, reading, and verbal ability). She also analyzed EEOR data from a number of

other cities (Philadelphia, Milwaukee, Washington, D.C., Cleveland, and Baltimore). The effects she found for quantity of schooling in the inner city schools were positive though smaller than their standard errors for all three dependent variables. The effect for quantity of schooling on achievement in the suburban schools was negative though nonsignificant ($p \geq .05$). Karweit (1976a) also performed similar analysis on a number of other sets of data without finding effects anywhere near as large as Wiley's.

Schmidt (1981) assessed the relationship between achievement and the number of hours of high school instruction in six curricular areas controlling for ability and student background. He used a national sample of 9,195 students in 725 schools from the graduating class of 1972. The data was collected as part of the National Longitudinal Study (NLS). The achievement measures were a vocabulary test of synonyms, a reading comprehension test, and a mathematics test. The ability measures consisted of a test of associative memory, a test of inductive reasoning and a test of perceptual speed. All six ability and achievement tests were developed by the NLS. The background variables consisted of sex, race (white, nonwhite) and a composite SES measure created from parent education, income, father's occupation and the possession of certain household items. The quantity of schooling in six subject areas (science, social studies, foreign language, English, math,

and fine arts) was computed for each subject matter using the number of semesters taken in each subject matter and the instructional time received during a semester. ACT test battery scores were available for a subsample of 1,421 of the students and the English, social studies, math and science subscales were used as additional measures of achievement for the subsample of students that had taken this battery.

Schmidt found that quantity of schooling had a clear positive effect on achievement controlling for ability and student background. This was true for both NLS and ACT achievement measures. As one might expect the relationship was strongest for time spent in classes closely related to the test material. The one exception was a strong relationship between foreign language exposure and all the achievement measures. Schmidt hypothesizes that ability was not adequately controlled for by the measures used, and time in foreign language classes was acting as a proxy for ability since college bound students tend to take more classes in foreign languages. The effects ranged from about two to four percentage points for every 100 hours of instruction (approximately one semester). The largest effects were for math achievement.

The divergent findings of Karweit as opposed to Wiley and Schmidt are somewhat puzzling given their similar methodology. Though Wiley's sample size of forty schools is relatively small, the size of the effects for quantity

of time were at least two and as much as three and a half times larger their standard errors. Karweit's Detroit sample of thirty schools was quite small and the absence of significant effects for quantity of schooling in that study could be explained by sampling error. This however was not the case for the study using a number of other large metropolitan areas. It is interesting that the same pattern of results was found for Detroit as the other metropolitan areas in terms of the sign of the relationship between quantity of schooling and learning in the inner city as opposed to the suburbs. Quantity of schooling at least as it was measured by Karweit and Wiley tends to be positively related to learning in the relatively low SES inner city while negatively related to achievement in the relatively high SES suburbs. This suggests that increased time in school may be an important factor for children in the inner city schools who are less likely to get exposure to the skills and content contained in school curriculum at home than their suburban counterparts. Schmidt's findings on this question were quite different than Karweit's and Wiley's. Schmidt examined the impact of quantity of schooling on achievement for students from six types of schools. Schools with high percentage of minority and low income students versus other schools in three size categories, less than 300 students, 300 to 600 students and above 600 students. The pattern of results he found was mixed with a strong tendency for school size to interact

w
o
a
g
r
n

with SES in terms of the impact of quantity of schooling on achievement. In some categories, for some types of achievement, quantity of schooling did seem to have a greater impact on low SES students. In other cases, the most striking being the impact of quantity of schooling in mathematics on mathematics achievement, the impact was much greater for students from schools with a high percentage of low SES students.

Schmidt's study differed from Wiley's and Karweit's among other things in the way time was measured. He used measures of the time spent in specific subject areas as opposed to the total time in school. This better match between content of exposure and the achievement measures used might explain why he found significant effect for quantity of schooling while Karweit did not.

ENGAGEMENT RATES

This section reviews a number of studies on engagement rates or the proportion of allocated instructional time in school a student actually spends attempting to learn. Common sense suggests that it is only during the portion of allocated time a student spends on task that learning can occur. There is also evidence that there is a stronger relationship between engaged time and achievement than between allocated time and achievement (Borg, 1980). Despite this, the time allocated to various types of instruction in schools is an important variable in

educational research for a number of reasons. First, it places a ceiling on the total amount of engaged time a student can receive. Secondly, allocated time can be directly affected by policy changes while engaged time is more directly under student control and can only be indirectly affected by teacher actions. Thirdly, as will be shown in this section, naturally occurring variations in allocated time seem to have at least as great an impact on engaged time as naturally occurring variations in engagement rates.

Karweit and Slavin (1981) collected data on the scheduled, instructional and engaged time in mathematics for the classrooms of twelve teachers. Six students in each classroom were observed for ten days. The students' activities were recorded every thirty seconds during math instruction. The mathematics computations, concepts and application subscales of the Comprehensive Test of Basic Skills (CBTS) were used as a pre and post measure of math ability. The means, standard deviations and correlations among the variables for lower and upper elementary classes in the study are given in Table 2 and Table 3.

The high intercorrelation among the three measures of time (scheduled, actual, and engaged), as well as the equivalence of their relationships to the achievement measure suggest collecting data on engaged time or even allotted time may not be necessary in time on task research. Scheduled time which is generally much cheaper and easier

Table 2

Lower Elementary (grades 2 and 3) N=33

Post Test 63.7(17.9)

Pre Test .91 62.3(17.6)

Scheduled .30 .23 94.1(11.5)

Instruct. .37 .29 .91 80.1(10.8)

Eng. Min. .42 .30 .87 .90 73.8(11.8)

Eng. Rate .37 .22 .19 .42 .64 .78(.07)

Table 3

Upper Elementary (grades 4 and 5) N=62

Post Test 56.6(19.5)

Pre Test .89 51.2(21.3)

Scheduled .41 .45 97.8(14.6)

Instruct. .36 .39 .97 84.9(14.5)

Eng. Min. .42 .43 .85 .89 75.9(14.6)

Eng. Rate .19 .15 .15 .19 .62 .78(.08)

to ob
in the
betwe
sched
time a
is a f
and en
schedu
measur
applic
Slavin
teacher
ranged
deviati

Ka

between
upper a
Evaluat
be disc
that in
engagem
found a

Al

reading
student
Study us
Six stu

to obtain seems to be an acceptable alternative at least in this small scale study for estimating the relationship between time and learning. It is true, however, that scheduled time is a substantial over-estimate of the actual time a student spends engaged learning a topic. If there is a fixed constant for the difference between scheduled and engaged time that is fairly stable across teachers, scheduled time could provide an acceptable substitute measure for the absolute amount of engaged time after applying a correction factor. The data from Karweit and Slavin suggest this is not the case. Across the twelve teachers the ratio of engaged time to scheduled time ranged from .42 to .81 with a mean of .67 and a standard deviation of .11.

Karweit and Slavin also found positive correlations between engagement rate and the three time measures in both upper and lower elementary classes. The Beginning Teacher Evaluation Study (BTES) also found similar results as will be discussed below. This is encouraging in suggesting that increasing instructional time may not lead to lower engagement rates at least within the range of variation found among the classrooms in this small scale study.

Allocated time, engaged time, and engagement rates in reading and mathematics among second and fifth grade students were recorded in the Beginning Teacher Evaluation Study using a procedure similar to Karweit and Slavin's. Six students within each class with average ability were

obse

simil

sugge

alloc

relat

As sta

time i

ment r

found

is con

Bo

rates c

observe

the stu

Borg al

who can

accurate

learning

of class

listened

Approxim

were rel

students

This sug

rates mi

given tha

ing atten

observed by a BTES staff member. The results were quite similar to those of Karweit and Slavin's. These results suggest that within the naturally occurring variation in allocated time, there is a zero to slightly positive relationship between allocated time and engagement rates. As stated above, this suggests that increasing allocated time in math and reading does not result in lower engagement rates due to fatigue or boredom. The BTES Study also found engagement rates ranging from about .70 to .75 which is consistent with the findings of Karweit and Slavin.

Borg (1980) reviewed a number of studies on engagement rates done in the 1920's and 1930's. The engagement rates observed then were somewhat higher than those observed in the studies discussed above, ranging from .80 to .98. Borg also discusses the question of how well an observer who can only assess outward signs of attention can accurately assess whether a pupil is actually engaged in learning. He cites Bloom (1976) who made sound recordings of classroom activities and asked students while they listened to them what their thoughts had been at that time. Approximately 65% of the students' thoughts during lecture were related to the lecture topic, while 55% of the students' thoughts during discussion were on discussion. This suggests observer's reports of engagement time and rates might be somewhat inflated. This is not surprising given that a student could seem to an observer to be paying attention while he or she was actually daydreaming.

nati
to h
enga
ment
95%.
as m
the
for s
when
Obvio
can h
exper
cated
great
those
data f
found
time i
arts,
at lea
as cla
area.
amount
much ti
of allo
done on

It was stated in the beginning of this section that naturally occurring variations in allocated time seemed to have as great if not greater impact on engaged time as engagement rates. In the studies discussed above, engagement rates have been found as low as 55% and as high as 98%. This implies that it would take approximately twice as much allocated time in a given instructional area when the engagement rate was at the very low end of the range for students to receive a given amount of engaged time as when the engagement rate was at the high end of the range. Obviously changes in engagement rates of this magnitude can have a large impact on the engaged time students experience in different topics. The variability of allocated time in various subject matters suggests it has even greater impact on the engaged time students receive in those subject matters. An analysis of the time allocation data from the six classrooms this study is based upon found that classes with the highest amount of allocated time in each of five major subject matter areas (language arts, reading, math, social studies, and science) spent at least twice as much time in that subject matter area as classes with the lowest amount of time in that particular area. In the case of science the class with the most amount of allocated time spent approximately 50 times as much time in science as the class with the least amount of allocated time in science. Although this study was done on a small number of classes, the results are

consi
(Borg
recei
matte
recei
stude
in th
time.

there
sched
mathe
They
able
In ad
acros
study
betwe
that
least
Engag
55% u
that
Altho
of en
of al
class

consistent with what has been found by other researchers (Borg, 1980; Mann, 1928). Even if classes where students received the most amount of instruction in a given subject matter area had the lowest engagement rates, they would receive more time on task in that subject matter area than students in classes with the least amount of instruction in that area even if they were engaged nearly 100% of the time.

In summary, Karweit and Slavin's study suggests that there is a near perfect linear relationship between scheduled time, allocated time and engaged time at least in mathematics instruction at the elementary school level. They also found however that scheduled time is a considerable overestimate of the allocated time in math instruction. In addition, the extent of overestimation is not consistent across classes. Karweit and Slavin as well as the BTES study found a zero to slightly positive correlation between engagement rates and allocated time suggesting that it is possible to increase instructional time at least to some extent without lowering student attention. Engagement rates in the studies reviewed have ranged from 55% up to nearly 100%. The work of Bloom (1976) indicates that engagement rates may be somewhat overestimated. Although there is some evidence that there is a wide range of engagement rates across classes, the even wider range of allocated time to various curricular areas across classes suggests that allocated time has as large, if not

large

a gi

THE

teach

resea

inves

been a

Rosens

studie

review

H

accura

focus

allocat

the stu

behavio

questio

general

informat

tion fro

Of

a clear

other so

teachers

classroom

larger, an effect on engaged time as engagement rates in a given subject matter area.

THE RELIABILITY OF TEACHER RECORDED DATA

This section reviews the research on the ability of teachers to provide accurate information for use in research. This topic is rarely the major focus of an investigation and what research has been done has generally been a by-product of research on other topics (Hook and Rosenshine, 1979). One study and a review of 11 other studies were found that relate to this question. The review article will be discussed first.

Hook and Rosenshine reviewed 11 studies of the accuracy of teacher reports. Although none of the studies focused on the ability of teachers to record student time allocation, their findings seem relevant. They grouped the studies into those of teacher reports of specific behaviors, those of scales formed from items in teacher questionnaires, and those of teacher reports grouped into general traits such as open versus traditional. The information provided by the teacher was related to information from an observer or students.

Of the six studies of specific behaviors, none found a clear relationship between teacher reports and the other source of information. Although this suggests that teachers cannot provide accurate information on specific classroom activities, other factors may in part explain

this

beha

and

cann

numb

1978

cong

may :

error

obser

for s

assum

two s

teach

was f

more s

accura

assess

classr

the ac

One wo

activi

less a

A

outsid

this finding. Studies of the generalizability of teacher behavior have found that the variability among occasions and raters is so large for certain behaviors that they cannot be measured reliably without the use of large numbers of raters and occasions (Erlich and Shavelson, 1978), (Shavelson and Dempsey-Atwood, 1976). The lack of congruence between teacher self reports and observations may in part be due to error in observations as well as error in teacher self reports.

The correspondence between teacher reports and observations was better for scales and dimensions than for specific behaviors. This is what one would expect assuming the error in teacher reports was random. In the two studies relating teacher reports of their general teaching style with observer ratings a strong relationship was found.

The results of this review are not surprising. The more specific the information a teacher provides, the less accurate it is likely to be. Although these studies assessed the reliability of teacher self reports of their classroom behavior, the findings may well generalize to the accuracy of teacher recorded time allocation data. One would expect that the finer and more specific the activity categories used in a time allocation study, the less accurate teacher logs would be.

A comparison of time allocation data collected using outside observers and teacher logs was done as part of the

Beginning Teacher Evaluation Study (BTES) (Fisher, Filby, Marliave, Cahen, Dishaw, Moore, and Berliner, 1978). The sample consisted of 25 second grade and 22 fifth grade classrooms. Three boys and three girls were selected in each class as target students. The teachers in each class kept daily logs of the time spent by each of the target students in specific content categories within the areas of reading and mathematics. On one day each week, a trained observer recorded each of the target students' activities and the times they occurred as well as error and engagement rates. Achievement data was also collected at four time points.

The BTES Study found that although miscategorization from the teacher logs did occur, there was in general a good match between the observations and the logs. The correlations between observation time and log time were reasonably high. In second grade they ranged from .44 to .95 with a mean of .68 across the different activity categories. In fifth grade they ranged from .06 to .94 with a mean of .65. From their experience and the data they collected, the researchers felt that teachers tended on an individual basis to overestimate or underestimate the time in different categories. Using the results of comparing the mean observation and log time, correction factors were computed for each teacher's bias by forming ratios of the observation time over the log time for reading and math. These ranged from 0.717 to 1.643 for reading

and
at 1
allo

can
their
teach
time
stude
math.
track
on th
the B
to ev
For t
bias
inves

and 0.422 to 2.727 for math. These results suggest that at least some teachers have a fair amount of bias in the allocated time they record.

The results of the BTES Study suggests that teachers can in fact collect reasonably accurate data on the time their students spend in different activities, though some teachers tend to overestimate or underestimate allocated time. In the BTES Study data was collected on only six students per class and just in the areas of reading and math. In the present study the ability of teachers to keep track of the activities of all the students in the class on the full range of subjects taught was assessed. Unlike the BTES study, the major focus of this dissertation was to evaluate teachers as a source of time allocation data. For this reason, the nature and extent of the error and bias in teacher recorded logs of student activities was investigated in much greater detail.

this

descri

data

ing d

detail

T

collec

first

per day

This is

research

the lev

Harnisch

to be co

Par

evaluati

student

transcrip

of a subs

reliabili

CHAPTER THREE

THE DESIGN OF THE STUDY

This chapter begins with a discussion of the questions this study attempts to answer. This is followed by a description of the six classrooms on which time allocation data were collected. Since the major focus is on evaluating data collection procedures, these will be described in detail next.

Two basic approaches were used to assess errors in collecting time allocation data in the classroom. The first approach assessed error at the level of total time per day for a given student in a given activity or pursuit. This is generally the lowest level at which time allocation research is done. The second approach assesses error at the level of the individual pupil pursuit as defined by Harnischfenger and Wiley (1976). This level was chosen to be consistent with Harnischfenger and Wiley's model.

Part of the data collection procedure this study is evaluating consists of coding written descriptions of student activities in the form of teacher logs or observer transcripts into pupil pursuit records. Multiple codings of a subset of the observer notes were used to assess the reliability of coders.

RES

this

ing

The

stru

stud

reco

writ

measu

or ac

proba

obtain

focus

studen

on tea

activi

expens

a full

these c

introdu

Th

and und

substitu

allocati

procedur

RESEARCH QUESTIONS

As was stated in chapter one, the major purpose of this dissertation is to evaluate two approaches to collecting classroom time allocation data on several dimensions. The first is the use of observers recording in the form of structured field notes the activities of individual students. The second is the use of classroom teachers recording the activities of their students in the form of written logs.

There is probably no way to obtain perfectly accurate measures of the time students spend in different pursuits or activities. The use of a full time observer can probably provide as accurate a description as can be obtained of student classroom activities. An observer can focus all of his/her attention on the task of recording student activities while a teacher's main focus must be on teaching. The use of teachers to record student activities, however, is likely to be considerably less expensive in that it eliminates the cost of the salary of a full time observer. The use of teachers to collect these data is also likely to be less disruptive than the introduction of an outside observer to the classroom.

The major focus of this dissertation is to what extent and under what circumstances can teacher logs be used as a substitute for outside observers as a data source for time allocation studies. In addition how might log keeping procedures and training methods be improved to increase the

acc

the

des

reco

sing

dine

zatt

acti

incl

a new

chang

very

It is

proce

7

what i

and so

differ

subjec

cussed

measuri

it's ef

among s

teacher

to be li

accuracy of the information provided by teachers.

The second question this dissertation addresses is the reliability of a set of procedures for coding written descriptions of student activities into pupil pursuit records. These pupil pursuit records indicate for a single student his/her activity coded on the three major dimensions of the Harnischfenger and Wiley model (subject matter, grouping composition and whether or not the activity was supervised by the teacher). The records include the beginning and ending times of the activity, and a new record was started when the students' activity changed according to any one of the dimensions.

The coding of the teacher logs and observations is very time consuming and tedious using this coding process. It is important to determine how reliable the coding process is and whether multiple coders are necessary.

The final question this dissertation addresses is what is the best strategy for sampling classrooms, students and school days within classrooms. If there are extreme differences in the time students receive in different subjects, as is suggested by the previous research discussed in chapter two, a great deal of precision in measuring time allocation may not be necessary to estimate it's effects on learning. If there is little variability among students within a class, as would be the case if the teacher mainly used whole group instruction, there seems to be little use in observing all or most of the students

in a class. On the other hand, if the teacher used a large amount of individualized instruction resulting in large differences among students within the class in the time devoted to different subjects, observing a large number of individual students would be important. As with students, if there is substantial variation among days in the school year within classes in the time spent in different subjects, it would be necessary to obtain data from a large number of days to get accurate estimates. If there was little variation among days, it would be necessary to obtain data on only a few days in each classroom to obtain reasonably accurate information.

SAMPLE

Time allocation data were collected in six elementary classrooms from the greater Lansing, Michigan area. The sample included classrooms from inner-city Lansing as well as rural and suburban districts around Lansing. There were two second grade classes, two third grades, a fifth grade and a team taught fourth and fifth grade double classroom.

DATA COLLECTION PROCEDURES

Classroom time data were collected during a three month period in the spring of 1978. During this period each teacher in the study kept daily logs of the classroom

act
were
taki
incl
stud
purp
prov

in th
class
the f
descr
indiv
beginn
transc
the ac
append

B
activi
appendi
observa
allow f
coders.

ESTIMATI

In
importan

activities of each student in the class. The teachers were asked to record the logs while the activities were taking place whenever this was possible. The logs included the beginning and ending time of an activity, student group, lesson content and materials, instructional purpose, and instructional strategy. An example is provided in appendix A.

On eight days in three of the classrooms and nine days in the other three classrooms, an observer recorded the classroom activities of each student in the classroom in the form of structured field notes. These included descriptions of the activities of each student group or individual, those students making up the group, and the beginning and ending time of the activity. An example of a transcribed version of a set of these notes describing the activities of a class for a day is provided in appendix A.

Both the written logs and observations of the students' activities were then coded using the scheme presented in appendix B. For two classrooms on four days each, the observations were coded by two separate individuals to allow for estimating the reliability of the individual coders.

ESTIMATION OF VARIANCE COMPONENTS

In order to help give a perspective as to the importance of the error introduced by various methods of

co
of
obs
exi
and
uni
wer
scie

comp
each
in s
eight
desig
rando
neste
other
thumb
devel
squar
the v
squar
progr
the c
diffe
measu

collecting classroom time allocation information, estimates of the variation among classrooms, measures (logs versus observations), students and days within classrooms and the existing interactions for five major subject matter areas, and the transitions between lessons were computed. The unit used was minutes per day, and the subject matters were language arts, reading, math, science and social science.

To greatly simplify the estimation of these variance components, days and students were randomly dropped from each classroom to obtain a balanced design. This resulted in six classrooms, two measures (observations and logs), eight days and 13 students within each classroom. In this design, classes, students and days were conceptualized as random and measures as fixed. Students and days were nested in classrooms and measures was crossed with the other three dimensions in the design. A set of rules of thumb developed by Millman and Glass (1967) were used to develop computational formulas for computing the mean squares and their expected values. Unbiased estimates of the variance components were obtained from the mean squares via algebraic manipulation. These formulas were programmed in FORTRAN by the author in order to perform the computations.

The reliability of measures of time allocation using different numbers of students, days and the two types of measures were computed using generalizability theory

(C

in

of

PRO

diff

the

stud

indi

proce

and o

diffe

The me

be des

logs a

for mu

P

contain

day wer

the num

student

pursuit

ject mat

social s

subgroup

(Cronbach, Gleser, Nanda, and Rajaratnam, 1972). This information can help researchers designing future studies of time allocation.

PROCEDURES WITH TIME PER DAY PER STUDENT AS UNIT

As was stated above, the error introduced by different approaches to collecting time allocation data in the classroom was assessed at the level of time per day per student (marginals for pupil pursuits) and at the level of individual pursuit records. This section describes the procedures at the first level.

At this level both the consistency between the logs and observations as well as the consistency among different coders coding the same observations was assessed. The methods used were generally parallel. The method will be described in terms of assessing the consistency of the logs and observations noting differences in the procedures for multiple coders where they exist.

Parallel data sets coded from logs and observations containing each student's pursuits in each class on each day were formed. The data sets were aggregated to obtain the number of minutes in pursuit categories for each student in each class on each day. The time in twelve pursuit categories was computed. These included five subject matters, language arts, reading, math, science, and social studies; three grouping strategies, whole group, subgroup and individual instruction; teacher supervised

a

t

ti

av

th

st

pr

es

ind

obs

tior

for

exte

out.

day

in a

the

days

and t

of th

show

or th

other

absolu

tude t

and unsupervised instruction, as well as the time in transition between lessons and mixed seatwork.

The average difference (observation time minus log time) was computed for each of the twelve categories and averaged across students and days in each classroom. For the multiple codings this was done across the days and students each pair of coders coded. These mean deviations provided a measure of the averaged difference in the time estimates for the different data collection methods.

For various students and or days, the logs may indicate more time in a given activity category than the observations, while for other students or days the observations might indicate more time. This could also be true for multiple codings of the same observations. To the extent this happened, the deviations would tend to cancel out. To provide a measure of the average difference per day per student between logs and observation time measures in a category, absolute values of the difference between the logs and observations were computed and averaged across days and students. A comparison of the mean difference and the mean absolute value difference provides a measure of the extent either the logs or observations consistently show more time in a pursuit category. If either the logs or the observations show more time in a category than the other across all days and students in a class, the mean absolute value of the difference will be equal in magnitude to the mean difference in that category. To the

exte
show
the
than

pret
diff
was
mult
indi
aver
stud

MEIR

data
Sinc
was
disc
achi

ther
reco
obse

extent it varies as to whether the observations or logs show more time in a pursuit category, the magnitude of the mean absolute value of the difference will be higher than the mean difference.

In order to provide a frame of reference for interpreting the mean differences and mean absolute value differences in the categories, the mean observation time was also computed for each category. In the case of multiple codings of the same observations, the time indicated in a pursuit category for a student on a day was averaged and these values were averaged across days and students.

METHOD USED TO ASSESS ERROR AT THE PURSUIT LEVEL

An attempt was made to evaluate the teacher logs as a data collection procedure at the individual pursuit level. Since this is the level at which the data is collected, it was felt that a better understanding of the types of discrepancies between the logs and observations could be achieved at this level.

Due to discrepancies in the logs and observations, there was not always a one to one match between pursuit records. A coding procedure was used to match log and observation pursuit records.

CODING

Gi

and log

and 23,

Four da

there w

class o

activit

of the

represe

This pr

Th

observa

extent

do this

observa

created

from th

was add

element

In

be a mo

than th

much sm

This re

breaking

pieces

CODING PROCEDURES

Given the number of pursuit records in the observation and log data sets (approximately 47,000 in the observations and 23,000 in the logs) a sample was selected for coding. Four days from each class were randomly sampled. Since there was a large amount of similarity among students in a class on a given day in their pursuit records due to group activities, three or four students in each class on each of the four sampled days were selected that tended to represent groups of children with similar pursuit records. This procedure resulted in 76 student-days being coded.

The purpose of the coding procedure was to match the observation and log pursuit records so that the type and extent of differences between them could be studied. To do this, the log pursuit records were mapped onto the observation pursuit records. That is, a coded record was created for each observation record and the information from the appropriate log pursuit record or portion of it was added to the coded record. Table 4 gives the data elements on each coded record.

In general, due to the fact the observations tended to be a more detailed description of the student activities than the logs, each observation pursuit record included a much smaller time interval than a log pursuit record. This resulted in the coding process to a large extent breaking up log pursuit records and assigning separate pieces to different observation pursuit records. The

cod

the

puz

to

Fir

sel

of

cla

cat

who

cla

a cl

from

were

coded records contained an ID number for the log pursuit the log portion came from if it was a piece of a larger log pursuit. This was so that it could be tied back to the total pursuit from which it came.

Table 4

Pursuit Record Data Elements

1. student ID code
2. class ID code
3. date
4. starting time of observation pursuit
5. total time of observation pursuit
6. subject matter code of observation pursuit
7. grouping code of observation pursuit
8. supervision code of observation pursuit
9. total time of log pursuit
10. subject matter code of log pursuit
11. grouping code of log pursuit
12. supervision code of observation pursuit
13. log pursuit ID if log pursuit divided

The coding process was made up of the following steps. First, the days to be coded for each class were randomly selected. The pursuit records were aggregated to the level of the total time for the day for each student in the class in the subject matters, grouping and supervision categories. The information was used to group students who represented different time allocation patterns in the class. There were generally three or four such patterns in a class on a given day. A student was randomly selected from each group and the pursuit records for these students were coded.

Listings of the log and observation pursuits on the

selected student-days were made. These were used to map the log pursuits onto the observation pursuits. The information from the observation pursuits was pulled off via a FORTRAN program while the matching log pursuit information was key-entered under control of the FORTRAN program.

In general the coding process of matching the log pursuits with the observation pursuits was quite straightforward. In most cases the pursuits matched in an obvious way one to one or as stated above, the log pursuits were broken up and portions of a single log pursuit were matched with a number of observation pursuits and an ID number assigned to them so that they could be tied back together if so desired during analysis. There were a few instances where an observation pursuit record included a large block of time where the log pursuit file had the same block of time broken into several pursuits. Since the coding procedure mapped the log pursuits onto the observation pursuits, there was no simple solution to this situation. There were multiple log pursuit records describing different time portions of the same observation pursuit record. In general the largest log pursuit record was matched with the observation pursuit record. Each occurrence of this situation along with how it was resolved is documented in appendix C.

Since the coding procedure used to match log and observation pursuit records could introduce error that

wou
obs
sel
of

THE

dif
rec
Sec
cat
jec
bot
est
agg
ana

rec
the
and
the
com
tim
as
che
err
rec

would be mistaken for real differences between the log and observation pursuits, two student days were randomly selected and coded by two individuals to provide a measure of the reliability of the coding procedure.

THE ANALYSIS OF CODED PURSUIT RECORDS

The pursuit records from logs and observations could differ in two ways. First they could differ in their recording of the beginning and ending time of the pursuit. Secondly, they could differ in how the activity was categorized on one or more of the three dimensions, subject matter, grouping and teacher supervision. Although both types of discrepancies result in differences in time estimates between the log and observation data when aggregated, these two types of discrepancies will be analyzed separately.

Both the teachers and the observers were instructed to record the beginning and ending times of activities using the wall clock in the classroom. The coding of the logs and observations were checked by multiple individuals and the coded pursuit records were verified keypunched. A computer check was also done to assure that the starting time of the next pursuit record for a student was the same as the ending time of the last pursuit record. These checks made it unlikely that any substantial amount of error was introduced into the log or observation pursuit records in terms of beginning and ending times during

codi

the

acti

keep

whil

had

char

at t

reco

less

on r

cate

proc

tior

teac

same

days

of t

tran

The

teac

cate

due

coding and data entry.

It seems reasonable to assume that discrepancies in the beginning and ending times of pupil pursuits or activities were mainly due to the teachers' inability to keep track of the exact times student activities changed while teaching. In many instances the teachers probably had to reconstruct or guess at the time that activities changed because they were too busy to look at the clock at that time. It is true that the observer could also record the times incorrectly, but this seems considerably less likely given the fact the observer's main focus was on recording student activities.

Differences in how a given student activity was categorized resulted from the unreliability of the coding process and differences and ambiguities in the descriptions of the activities contained in the observer notes and teacher logs. As described above, multiple codings of the same observer notes were done for two classrooms on four days each. This should provide data on the unreliability of the coding process. Examples of teacher logs and transcribed observer notes are contained in appendix A. The far greater detail of the observer notes as compared to teacher logs would suggest that discrepancies in the categorization of activities would for the most part be due to ambiguities in the teacher logs.

THE ANALYSIS OF THE CATEGORIZATION OF PURSUITS

Discrepancies in the categorization of student activities was assessed as follows. A crosstabulation like table was constructed with one of the dimensions being the log categorization and the other being the observation categorization. The following statistics were included for each cell of the table; the number of pursuits in that cell, the average number of minutes of the pursuits (from the observations), and the proportion of pursuits in that cell of the total number of pursuits with that given observation coding. Tables were constructed for activity type (the five subject matters, transitions between lessons, and seatwork), group type and supervision.

The diagonal cells in this table represent those pursuits where the categorization of pursuit by the logs and observations was consistent. Each off-diagonal cell represents a certain type of discrepancy between the log and observation coding of the activity. It is possible to evaluate the extent and nature of the differences in the log and observation codings of the student activities from these tables.

A log-linear model (Bock, 1975) was fit to test whether there were statistically significant differences among classrooms in the proportion of pursuits in the cells of the table described above. The differences among classrooms was found to be significant beyond the $p \leq .001$ level for activity type as well as group and supervision

code. For this reason tables for each classroom as well as a table of the combined results are presented and discussed.

ANALYSIS OF ERRORS IN PURSUIT LENGTH

A measurement model described by Schmidt (1981) was used to assess the consistency of the length of log and observation pursuit records. As has been discussed above, the assumption has been made that the observation data is more accurate than the logs. The purpose of comparing the use of teacher logs and observer transcripts as methods of collecting classroom time allocation information is to see how well the information collected from logs approximates that from observations. For this reason, the observation time has been conceptualized as a true score in this model and the log time as an observed score.

The advantage of Schmidt's model over the classical true score model in this situation is that it allows for fixed bias and random bias correlated with the true score as well as random error independent of the true score. Fixed bias in this situation would be the consistent tendency for a teacher to over or under estimate the time in an activity category. Random bias correlated with the true score in this study would be the tendency for the bias in the teachers' estimates of the students' time in an activity to change with the magnitude of the true time in the activity. In addition the model includes a random

error term uncorrelated with the true score with an expected value of zero.

Schmidt's model is expressed as follows:

$$X = A + B\xi + \epsilon \quad (1)$$

where X is the observed score (log time), ξ is the true score (observation time) and ϵ is the random residual term in the model. A is the intercept and B is the slope of the linear relationship between X and ξ . It is also assumed that the covariance of ξ and ϵ is zero. Under this model measurement error denoted by e is expressed as X minus ξ .

Given the model of (1), and the definition of measurement error as the difference between observed and true scores, measurement error (e) can be expressed as follows:

$$e = A + (B-1)\xi + \epsilon$$

As one can see, measurement error in this model is made up of three distinct components. The first, specified by A is a fixed bias. The second which is a function of both B and ξ , represents a random bias associated with the magnitude of the true score. The final component ϵ , is random error independent of the true score.

The model was fit for each class separately as it was felt that both bias and random error in keeping track of pursuit length would vary among the teachers. The model was also fit for the five subject matter categories,

transitions between lessons, and seatwork as well as the grouping and supervision categories. For the pursuit categories, the model was fit for only those pursuits where there was agreement between the logs and observations to the pursuit category. Both log and observation measures of pursuit length are random variables. For this reason maximum likelihood was used to estimate the parameters and provide correct asymptotic standard errors.

SUMMARY

This chapter described the methods that assess using teacher logs as opposed to outside observers as a method of collecting classroom time allocation information within the framework of the Harnischfenger and Wiley model of school learning. Teacher logs and outside observer notes were collected from six classrooms for eight (three classrooms) or nine (three classrooms) days on each student's activities. These descriptions were coded into pursuit records as defined by the Harnischfenger and Wiley model. In the case of two classrooms for four days each, the descriptions provided by the observers were coded by two separate individuals.

Five approaches were used to assess the consistency of the information provided from the logs and observations and evaluate the data collection procedure as a whole. The expected values of the variability among classrooms and students and days within classrooms were computed for

the time in minutes spent in language arts, reading, math, science, social studies and transitions between lessons. This was done both to provide some perspective on the discrepancies found between activity time computed from logs and observations as well as help determine how many days and students a researcher would need to observe to obtain accurate activity time estimates for a classroom.

The consistency of multiple codings of the observations from the four days each in two classrooms was assessed by computing average differences, and average absolute value differences across students and days for each pair of coders in the five subject matters listed above, transitions between lessons, group type, and the presence or absences of teacher supervision after the pursuit records were aggregated to the time per category per student per day. A parallel analysis was done comparing the time coded from teacher logs and observer transcripts.

Since the descriptions of student activities were collected and coded at the level of the individual pursuit records, it was felt that the discrepancies between the logs and observations could be best understood at that level. A coding procedure was used to match a sample of the log and observation pursuit records. Two types of discrepancies could exist between pursuit records coded from logs and observations. First, they could differ in length. Second, they could differ in how the activity was

categorized. A measurement model developed by Schmidt (1981) was used to estimate random and fixed bias, as well as random error in the log pursuit length as compared with the observation pursuit length. Crosstabulation of how the pursuit records were categorized from logs and observations for all the classrooms as a whole as well as the individual classrooms, was done to assess the discrepancies in pursuit categorization.

CHAPTER FOUR

THE RESULTS

The results of the data analysis are presented in this chapter. Five approaches were used to evaluate the reliability and efficiency of the data collection methods used by the Language Arts Project to collect time allocation data in classrooms. The first approach evaluated the reliability of coders coding written observations into pupil pursuit records. The second approach estimated the variance components for different subject matters; 1) among classes, 2) among schools within classes, 3) among students within classes, and 4) among measures (logs and observations), as well as all the existing interactions. These variance components were then used to estimate the reliability for different data collection designs. The third approach assessed the consistency of measures of time in different pupil pursuits collected with teacher logs and observer notes at the level of student time per day. The fourth approach analyzed the discrepancies in the categorization of individual pupil pursuit records coded from logs and observations. The fifth approach estimated fixed and random bias as well as random error in teacher recorded pursuit length as compared with observer

recorded pursuit length.

ESTIMATION OF VARIANCE COMPONENTS

In this section, estimates of the variance components among classes, measures (logs and observations), students, days and the existing interactions of these factors are presented. These variance components are then used to estimate the reliability of four data collection designs using generalizability theory (Cronbach, et al.).

The variance components are presented in Table 5. They were estimated for language arts, reading, math, science, social studies and transitions between lessons. The sources of variance include measures (logs and observations), classes, days, students, and the following interactions: measures by class, students by day, measures by day, measures by student and measures by student by day. No other interactions existed due to the nesting of students and days within classes. The numbers in the table are the variance estimates while the numbers in parentheses are the square roots which are probably more useful in that their metric is minutes.

There were relatively small amounts of measures variance across the five subject matters and transitions between lessons. Measure variance is the variability between time measures from logs and observations averaged over students, classes and days. For math, science and transitions, there was no variation among measures. For

Table 5

Variance Components of Measures of Five Subjects and Transitions

	Language Arts	Reading	Math	Science	Social Studies	Tran- sitions
Measures	3.17 (1.78)	7.19 (2.68)	0.00 (0.00)	0.00 (0.00)	9.36 (3.06)	0.00 (0.00)
Classes	173.28 (13.16)	665.12 (25.79)	96.46 (9.82)	302.21 (17.38)	239.59 (15.48)	137.83 (11.73)
Days	807.77 (28.42)	423.11 (20.57)	158.00 (12.56)	243.95 (15.62)	265.89 (16.31)	41.05 (6.41)
Students	29.26 (5.41)	40.10 (6.33)	1.60 (1.26)	0.00 (0.00)	0.00 (0.00)	0.81 (0.90)
M x C	13.47 (3.67)	0.00 (0.00)	11.46 (3.39)	1.49 (1.22)	0.00 (0.00)	165.91 (12.88)
S x D	15.51 (3.94)	36.80 (6.07)	4.40 (2.10)	7.38 (2.72)	8.80 (3.00)	0.90 (0.95)
M x D	247.56 (15.73)	252.86 (15.74)	82.95 (9.11)	92.84 (9.64)	250.33 (15.82)	85.61 (9.25)
M x S	0.00 (0.00)	22.62 (4.76)	0.00 (0.00)	0.00 (0.00)	0.56 (0.74)	0.42 (0.65)
M x D x S	184.17 (13.57)	50.41 (7.10)	55.66 (7.46)	3.64 (1.91)	31.13 (5.58)	25.76 (5.08)

language arts, reading and social studies, the square roots of the variance components were 1.78, 2.68 and 3.06 minutes respectively. These results suggest that when the design is collapsed across classes, days and students, there is little or no variability between log and observation recorded time.

The variability among classes was relatively large when compared to the other variance components. For reading and science it was the largest source of variance. For transitions it was the second largest, while for math and social studies it was the third, and in language arts it was the fourth. The square root of the variance among classes ranged from 11.73 minutes in transitions to 25.79 minutes for reading. As will be discussed in the next section, variation among classes would be conceptualized as "true score" variance when a research is interested in the differences in time spent in a type of activity among classrooms. The fact that this source of variance is large suggests that differences among classes in the time spent in various activities can be measured reliably.

The variability among days within classes was also large when compared to the other variance components. It was the largest component for language arts, math and social studies, and the second largest for reading and science. Days was the fourth largest variance component for transitions. Apparently there are large day to day differences in the amount of time allocated in these five

subject matters within classes. Since day to day variation in allocated time to a subject matter would generally be considered error, increasing the number of days each class was observed would be an effective approach to increasing the reliability of time allocation measures.

Compared with days and classes, the variability among students within a class in the time spent in each of the six activity categories was quite small. In science and social science, this component was zero. Reading had the largest amount of variability among students of the six activities where the square root of the variance component was 6.33. Apparently students within the same class tend to receive pretty much the same amount of instructional time per subject matter.

The measure by class interaction was quite small for the five subject matters but relatively large for transitions. In reading and social studies it was zero, and for the other three subject matters the square root of this component was less than four minutes. It was however the largest component for transitions where the square root of the variance component was 12.88 minutes. The size of this component has important implications in terms of this study. It is the extent that there were differences among classrooms in the discrepancies between log and observation recorded time. If one assumes that the observation measures were accurate as has been done in this study, the class by measure interaction is a measure of the extent

teachers differ in the magnitude and/or sign of the error in their recording pursuit length averaged over students and days. These results suggest that the differences among teachers in the magnitude and direction of their errors in recording of student activities was generally small with the exception of transitions. As will be presented in a following section, there were in fact large differences both in sign and magnitude among teachers in the differences between log and observation recorded time for transitions.

The square root of the student by day interaction component ranged from just under a minute in transitions to just over six minutes in reading. It was the sixth largest of the nine components for all the activity categories with the exception of science, where it was the fourth largest.

The measure by day interaction component was fairly large across all the activity categories. It's square root ranged from 15.84 minutes in social studies to 9.11 minutes in math. This suggests that there are large day to day differences in the discrepancies between logs and observations within a class. Whole class and subgroup activities make up a large portion of a school day in most classrooms. The activity description recorded by the teacher or observer would be the same for all the students in the class or subgroup when whole class or subgroup activities were taking place. Errors and ambiguities in a

teacher's description of a student activity or it's length would occur for all the children in the class or subgroup if the whole class or subgroup participated in the activity. This could result in large discrepancies on a given day between log and observation time recorded for a given activity type. This might explain the large measure by day interaction. If the errors in the teacher logs were basically random, they would cancel out across days resulting in little variance between the log and observation measures averaged over days as was observed in this study.

Unlike the measures by day interaction, the measures by students interaction was small or nonexistent. This component was zero for language arts, math and science. The square root was under a minute for the other categories with the exception of reading, where it was 4.76 minutes. Again the fact that much of the student day is spent in group activities could explain the small student by measure interaction. Discrepancies between log and observation recorded time would be consistent across students in the group or in other words no student by measure interaction.

The measures by day by student interaction was moderate in size for the different activity categories. Since there was only one observation per cell in this design, this component includes the residual variance from all sources not included in the design. The square root of this component ranged in magnitude from 1.91 minutes in

science to 13.57 minutes in language arts. It is difficult to interpret this component, first because it is a three way interaction, and secondly because it also contains all other sources of variance not included in the design.

The way the variance was distributed across the nine components was fairly consistent for the six activity categories. The largest sources of variance were classes, days and the measure by day interaction with the exception of transitions, where the measures by class interaction was the largest source. There were relatively small amounts and in the case of science, math and transitions no variance among measures. This suggests that the differences between teacher logs and observer field notes in the time recorded in these activity categories tends to cancel out when averaged over days, students and classes.

Through the use of the variance components discussed above, the reliability or generalizability of time measures in a number of research designs was computed. For studies of differences among classes, the designs included all combinations of using logs or observations, measures on two and four days, and observing five and ten students. For studies of differences among students within a classroom, the designs included all combinations of using logs or observations and measures on two and four days. The results are presented in Table 6. The generalizability coefficients are to some unknown, though likely to some extent, underestimated for the designs using

Table 6

Generalizability Coefficients for Measuring Time Allocation

	Observations					
	Among Classes				Among Students	
	Days				Days	
	<u>Two</u>	<u>Students</u>		<u>Four</u>	<u>Two</u>	<u>Four</u>
	<u>Five</u>	<u>Ten</u>	<u>Five</u>	<u>Ten</u>		
Lang. Art	.29	.29	.44	.45	.05	.10
Reading	.74	.75	.85	.86	.14	.24
Math	.53	.54	.69	.70	.01	.03
Science	.71	.71	.83	.83	.00	.00
Soc. Stud.	.64	.64	.78	.78	.00	.00
Trans.	.86	.86	.92	.93	.02	.05
Logs						
Lang. Art	.23	.24	.37	.38	.04	.08
Reading	.64	.65	.77	.78	.09	.15
Math	.41	.42	.56	.57	.01	.02
Science	.64	.64	.78	.78	.00	.00
Soc. Stud.	.47	.47	.63	.63	.00	.00
Trans.	.37	.37	.41	.41	.01	.02

observers. The measure by day by student interaction was included in the estimate of the observed variance because this component contains the variance from all sources not included in the design. The actual variance of the three way interaction should not be in the observed variance due to the fact observers are considered accurate and differences between observations and logs are due to errors in the logs.

The results in Table 6 suggest that both increasing the number of days observed and using observers as opposed to teachers to collect data result in meaningful improvements in reliability. Increasing the number of students observed from five to ten when the focus is on measuring differences among classes made at most a hundredth of a point improvement in the reliability.

The reliability for assessing differences among classes was in general reasonably good. Although it was somewhat low for language arts and math in the designs presented and for transitions when logs were used, acceptable levels of reliability, e.g. .80 and above, could probably be achieved by collecting data on more days and/or using observations. This was not the case when the focus was on assessing differences among students within classrooms. For science and social science there was no variance among students within a class and hence differences cannot be measured at all. The variability among students was small for the other four categories. It is

likely that reasonably reliable data could be obtained for reading and possibly language arts if data was collected on a large number of days particularly if observers were used. It is unlikely that acceptable levels of reliability could be achieved for measuring math and transition time with any sort of reasonable design.

The use of observers as opposed to teachers had the largest impact for measuring transition time. For measuring differences among classes, it more than doubled the size of the coefficients. This was due mainly to the fact the class by measure interaction was the largest source of variance for transition time. For the other activities, there was generally an improvement of between .05 and .10 in the reliability when using observers as opposed to teachers. The improvement was roughly the same or slightly better when data was collected on four as opposed to two days. Given the costs of using outside observers, increasing the number of days observed and using teachers as opposed to observers may well be a more cost effective approach for achieving equally reliable data at least for these six activity categories.

RELIABILITY OF THE CODING PROCEDURES

The observer notes from classrooms one and three on four of the days observed were coded by two separate individuals. These coders were graduate students who for the most part were working towards their doctorate in

education. In classroom one, separate sets of coders coded two days each. In classroom three, one individual coded four days while two other individuals each coded two days. The double codings were used to assess the reliability of coders using the coding conventions developed by the Language Arts Project. These coding conventions are contained in appendix B.

The pursuit records produced from each set of codings were aggregated to the level of time per day per student for each of the five subject matter areas, transitions between lessons, seatwork, supervision and grouping categories. The time in each of the above categories on each student-day from the two codings was averaged. In addition, the difference and the absolute value of the difference was computed. These three statistics were then averaged across students and days coded for each pair of coders. The results for the two classes are presented in Table 7.

The average time for the two codings in each category was provided to give some perspective for interpreting the differences in the time in different categories coded by different coders. An average difference between coders of three minutes has a different meaning when the total time in the category was 10 minutes as opposed to 30 minutes. The ratio of the difference of the codings over the average of the two codings expressed as a percent is also provided in Table 7. The mean deviations provide a

Table 7

Comparison of Time Measures from Multiple Codings

Class 1					
Coder 1 vs Coder 2					
Category	Ave.	Dev.	Abs. Dev.	% of Ave.	% of Abs.
Lang. Arts	34.98	1.96	2.54	6	77
Reading	84.73	2.81	8.24	3	34
Math	34.15	7.56	9.69	22	78
Science	8.18	1.52	2.31	19	66
Soc. Studies	11.81	-0.88	0.88	7	100
Seatwork	5.11	-2.15	6.06	42	35
Transitions	47.50	-12.21	12.67	26	96
Whole Group	55.86	-11.44	19.98	20	57
Subgroup	12.34	23.77	24.27	192	98
Individual	151.64	-1.44	9.85	1	15
Supervised	91.82	-3.02	21.48	3	14
Unsupervised	120.11	-1.90	6.60	2	29

Table 7 (cont'd.)

Class 1

Coder 3 vs Coder 4

Category	Ave.	Dev.	Abs. Dev.	% of Ave.	% of Abs.
Lang. Arts	41.17	-9.90	11.23	24	88
Reading	53.63	-3.40	4.98	6	68
Math	22.50	0.50	0.92	2	54
Science	0.00	0.00	0.00	0	0
Soc. Studies	0.21	-0.42	0.42	200	100
Seatwork	21.29	10.54	10.54	50	100
Transitions	34.19	0.46	5.17	1	9
Whole Group	78.92	-11.17	11.17	14	100
Subgroup	4.92	-0.58	5.08	12	11
Individual	129.47	7.94	8.85	6	90
Supervised	94.37	-7.77	11.31	8	69
Unsupervised	12.15	4.13	9.46	4	44

Table 7 (cont'd.)

Class 3

Coder 5 vs Coder 6

Category	Ave.	Dev.	Abs. Dev.	% of Ave.	% of Abs.
Lang. Arts	30.83	3.90	3.94	13	99
Reading	79.18	-1.32	8.32	2	16
Math	24.68	0.48	0.48	2	100
Science	65.19	-0.42	0.42	1	100
Soc. Studies	30.48	-0.48	0.48	2	100
Seatwork	0.71	1.42	1.42	200	100
Transitions	30.36	-0.20	0.44	1	45
Whole Group	189.98	4.06	4.78	2	85
Subgroup	3.40	-1.40	1.40	41	100
Individual	38.46	-0.44	4.16	1	1
Supervised	208.05	5.38	5.54	3	97
Unsupervised	23.70	-3.16	3.44	13	92

Table 7 (cont'd.)

Class 3

Coder 5 vs Coder 7

Category	Ave.	Dev.	Abs. Dev.	% of Ave.	% of Abs.
Lang. Arts	32.34	-0.24	2.76	1	9
Reading	73.85	2.10	2.18	3	96
Math	27.91	3.38	3.38	12	100
Science	50.27	5.66	7.06	5	80
Soc. Studies	53.85	-2.46	2.90	5	85
Seatwork	2.03	0.14	0.94	7	15
Transitions	31.06	-1.64	5.68	5	29
Whole Group	118.58	13.08	16.44	6	80
Subgroup	5.03	-0.38	3.26	8	12
Individual	31.98	-3.80	16.04	12	24
Supervised	236.94	8.72	11.76	4	74
Unsupervised	18.70	0.08	3.28	0	2

measure of the difference between coders in the average amount of time in a category coded over two days and approximately 20 students. Since the deviation between coders in the time they code for a pursuit category on a day for a student can be positive or negative, they can cancel each other out when averaged over days and students. The average absolute value of the differences was computed to assess the extent this occurred. The extent the mean absolute difference between coders across days and students is greater than the mean difference between coders can be looked upon as a measure of the extent the differences between coders is not systematic or in a sense random. For example, if one coder consistently codes more time in language arts than the other coder across days and students, the difference in the time in language arts between them will always be positive or negative and the mean difference and the mean absolute difference would have the same magnitude. To the extent it varied as to which coder coded more time in language arts across students and days, the magnitude of the mean absolute time difference would be larger than the mean time difference. The ratio of the mean difference over the mean absolute value difference expressed as a percent is provided in Table 7.

For the five subject matter areas, the reliability of the coding seems in general to be quite good. In most cases the mean difference between coders was less than 10%

of the average time coded in a category. The subject matters where the discrepancies exceeded 10% of the average time in the subject matter differed between sets of coders suggesting that there was no particular subject matter that was difficult to code. In some cases the mean differences and absolute mean difference were equal or nearly equal suggesting there were systematic differences between coders while in other cases the absolute mean difference was substantially higher than the average difference. With the exception on social studies where there were systematic differences between each pair of coders, there did not seem to be a tendency across coders for the differences between them to be systematic or not systematic.

For three of the four pairs of coders, the discrepancy in the time categorized as mixed seatwork was relatively large compared to the amount of time coded as mixed seatwork. This suggests that it may be difficult to distinguish whether an activity is mixed seatwork or instruction in a given subject matter. There was also a large discrepancy between one set of coders in terms of the time spent in transitions between lessons. The fact the absolute value of the difference between coders and the difference between coders were almost equal means that there is a consistent tendency for one of the coders to code more time as transitions than the other. One other set of coders was almost in perfect agreement as to the time

spent in transitions. The other two sets of coders had fairly small (.46 and 1.64 minutes) average differences per student per day. The absolute mean differences were between five and six minutes for both, suggesting it varied as to which coder coded more time in subgroup activities across days and students.

As with the subject matter categories, the consistency of the coding of the group structure varied between sets of coders. The greatest inconsistencies were found in coding the subgroup category. Coders one and two differed by almost 24 minutes while the average of their two codings was just over 12 minutes suggesting one coded on the average almost 24 minutes while the other coded essentially no time in subgroup pursuits. The other three pairs of coders had somewhat smaller differences in subgroup time. The mean absolute value differences however were quite large when compared to the amount of time coded in subgroup activities. In general, the coding was more consistent for whole group and individual activities. Two of the sets of coders had mean differences and mean absolute value differences greater than 10% of the average time coded for whole group or class activities. One set of coders had a mean difference and a mean absolute difference of greater than 10% of the average time coded for individual activities. The other pairs of coders coding whole group and individual time were quite consistent when compared to the amount of time coded for these

categories.

The coders were also fairly consistent in terms of their coding of the time spent in activities supervised and unsupervised by the teacher. The one major exception was that one pair of coders had a mean absolute value difference for supervised time of 21.48 minutes where the average time per student per day supervised was 91.82 minutes. The deviation averaged across days and students was only 3.02 minutes suggesting that one coder did not consistently code more time supervised than the other.

In summary, the coding in general was fairly reliable. In most instances the average difference between coders per day per student in an activity category was less than 10% of the average time per student per day in that activity category. In addition, there did not seem to be any activity category upon which all four sets of coders had large differences.

There did seem to be some difference between pairs of coders in the extent of agreement between them, though this conclusion is tentative given the size of the data set. Coders 5 and 7 were the most consistent. Only in math and individual instruction was the average difference between them greater than 10% of the average time coded between them in the category and in those categories the average difference was only 12% of the average time coded between them. Coders 1 and 2 on the other hand had averaged differences greater than 10% of the average time

in six categories and these percentages ranged from 19 to 192.

Given the fact that these results are based on a relatively small data set, they are difficult to interpret. Each set of coders only coded two days, and a single difference in the categorization of an activity to which the whole class participated in could result in a large systematic difference between the coders in the time coded in two categories. This probably to some extent makes the differences between coders seem more systematic than they in fact are. What appears clear from these results is that differences between coders are to some extent (and quite possibly a large extent) due to nonsystematic differences rather than systematic differences in how they categorize written descriptions of student activities. If the assumption is made that these nonsystematic differences between coders are due to random errors in each coding and these errors have an expected value of zero, the reliability of the coding process would improve with the size of the data set coded as these random errors cancel out.

THE CONSISTENCY OF LOG AND OBSERVATION ESTIMATES

In this section the results of comparing time on task estimated from teacher logs and observer notes at the level of time per student per day are presented. Pursuit time records coded from logs and observations were each aggregated to produce data records with the total time for

each student in the study on each day. Both log and observation data was available. This was done for time in language arts, reading, mathematics, social studies, science, seatwork, transitions between lessons, time in whole class activities, activities in subgroups within the class, individual activities, teacher supervised activities, and unsupervised activities. The average time per day per student in each category coded from the logs and observations in each class are presented in Table 8. In addition, these tables contain the ratio of the difference between the log and observation time over the observation time expressed as a percentage to aid in the interpretation of the magnitude of the difference. The ratio of the difference between the observations and the logs over the absolute value of the average difference between the observations and the logs expressed as a percentage is also contained in Table 8. This statistic can be used to assess the extent the daily differences between time coded in a category from teacher logs was systematically higher or lower than the time coded from observations. These analyses are essentially the same as was done previously in the analyses of the differences among coders. Table 9 presents the average difference between log and observation time in each class for language arts, reading, math, science, social science and transitions as a percentage of the square root of the estimated variance among classes and students (see Table 5). The square roots of the

Table 8

Activity Time Measure Differences

		Class					
		1	2	3	4	5	6
Lang. Arts	Obs.	38.96	41.88	36.02	44.32	68.87	67.92
	Log	52.42	54.75	30.32	48.12	78.02	65.22
	Dev.	35%	31%	16%	9%	13%	4%
	Abs.	91%	44%	66%	22%	31%	15%
Reading	Obs.	67.58	37.54	77.82	82.54	21.48	41.74
	Log	70.34	37.75	78.91	93.45	22.86	47.25
	Dev.	4%	1%	1%	13%	6%	13%
	Abs.	13%	3%	10%	39%	8%	35%
Math	Obs.	35.18	29.83	29.23	63.31	34.02	37.99
	Log	39.39	34.31	27.30	47.11	27.53	40.20
	Dev.	12%	20%	7%	26%	19%	6%
	Abs.	66%	91%	32%	93%	58%	21%
Science	Obs.	2.32	9.86	48.51	9.33	1.69	14.34
	Log	3.21	11.25	51.35	12.07	6.02	0.00
	Dev.	7%	14%	6%	29%	256%	100%
	Abs.	36%	100%	51%	100%	100%	100%
Social Studies	Obs.	4.42	8.61	32.24	0.00	39.07	17.88
	Log	3.21	11.83	34.95	5.06	42.91	33.96
	Dev.	27%	37%	8%	-	10%	90%
	Abs.	74%	100%	51%	100%	17%	69%
Seat- work	Obs.	16.64	44.90	5.51	7.09	32.57	5.04
	Log	0.00	22.44	3.81	8.00	40.70	24.66
	Dev.	100%	50%	31%	12%	25%	389%
	Abs.	100%	59%	90%	6%	63%	92%
Tran- sitions	Obs.	37.80	44.50	30.15	15.72	19.68	47.01
	Log	42.44	50.26	7.46	30.33	16.52	13.71
	Dev.	12%	13%	75%	93%	16%	71%
	Abs.	52%	37%	100%	86%	25%	100%

Table 8 (Cont'd.)

		Class					
		1	2	3	4	5	6
Whole Group	Obs.	82.75	75.07	186.70	101.58	92.61	103.82
	Log	95.07	90.78	179.63	108.54	123.80	119.33
	Dev.	15%	21%	4%	7%	34%	15%
	Abs.	94%	53%	24%	18%	53%	53%
Sub-Group	Obs.	8.72	8.79	6.29	19.94	63.50	17.15
	Log	0.00	9.26	4.50	29.95	15.05	16.71
	Dev.	100%	5%	28%	50%	76%	3%
	Abs.	100%	9%	30%	45%	90%	3%
Individual	Obs.	134.42	121.62	53.66	152.61	92.53	113.98
	Log	133.94	103.33	66.99	104.52	73.77	122.02
	Dev.	0%	15%	19%	32%	20%	7%
	Abs.	3%	50%	47%	96%	57%	21%
Super-vised	Obs.	103.48	106.31	216.57	122.92	151.50	140.17
	Log	207.86	140.44	234.20	203.46	202.93	193.24
	Dev.	101%	32%	8%	66%	34%	38%
	Abs.	100%	96%	72%	90%	92%	94%
Not Super-vised	Obs.	116.44	82.90	26.51	142.21	84.44	89.94
	Log	21.14	23.61	11.15	98.51	58.14	22.46
	Dev.	82%	28%	43%	69%	69%	25%
	Abs.	100%	94%	85%	100%	98%	60%

Table 9

Ratio of Time Measure Differences Over Sources of Variance

		Class					
		1	2	3	4	5	6
Lang.	Class	133%	127%	56%	37%	90%	27%
Arts	Student	260%	249%	110%	74%	177%	52%
Read.	Class	13%	1%	5%	55%	6%	25%
	Student	37%	3%	15%	148%	19%	75%
Math	Class	34%	47%	15%	130%	52%	18%
	Student	-	-	-	-	-	-
Sci.	Class	1%	8%	17%	16%	26%	86%
	Student	-	-	-	-	-	-
Social	Class	7%	20%	36%	31%	24%	99%
Stud.	Student	-	-	-	-	-	-
Trans.	Class	38%	47%	186%	120%	26%	273%
itions	Student	374%	465%	1830%	1178%	255%	2685%

estimated variances can be thought of as roughly the average difference between the mean of all classes (or students within a class) and each given class (or student within a class). Since these are the differences in which researchers studying time on task are generally interested, the relationship between the magnitude of the error in teacher logs as compared with observer notes with these differences should be very useful for interpreting the seriousness of these errors. The metric for all the above statistics is minutes per day per student.

It should be remembered that the discrepancies between pursuit records produced from logs and observations are due both to coder unreliability and differences between the logs and observations. There was no good way to separate these sources of differences between the log and observation pursuit records. If one compares Table 7 containing the differences between multiple codings of the same observations with Table 8 containing the differences between logs and observations, the differences between the logs and observations were in general considerably larger than the differences between coders. This was true despite the fact the observation versus log means were based on eight days, while the multiple codings were based on only two days, giving a much greater opportunity for nonsystematic differences between observers and teachers as opposed to multiple coders to cancel each other out over days. This suggests that a major portion of the discrepancies

between the logs and observations are due to real differences in the information they contain rather than to coder unreliability in translating the written descriptions into pursuit records.

The discrepancies between logs and observations in the average time per day per student in language arts ranged from 4% to 35% of the average time in language arts computed from the observations across the six classrooms. There was also a substantial difference across classrooms in the extent the differences between the log and observation time was systematic. The ratio of the difference over absolute value difference between logs and observations ranged from 15% to 91%. As was stated above, one way to assess the seriousness of differences between the logs and observations is to compare their magnitude with the variability among classes and students within classes in language arts time. The ratio of the average difference between log and observation time in language arts over the square root of the estimated variance among classes and students within classes is presented in Table 9. These ratios ranged from 27% to 133% across classes for the square root of the variance among classes and 52% to 260% for the square root of the variance among students within classes. In two of the classes for class variability and four of the classes for student variability, the ratios were greater than 100%. It would seem reasonable to conclude in such cases, substituting teachers as a source of

time on task information for observers would be questionable if a researcher is interested in investigating differences among classes or students within classes. The fact that in four of the classes the average difference between log and observation time measures was less than half as large as the average absolute value difference suggests that teacher logs may be an acceptable source of data in large scale studies where the error in the data they provide would cancel out.

The consistency between the logs and observations in terms of the time spent in reading was substantially better than language arts. The difference between log and observation time measures ranged from 1% to 13% of the time in reading measured from observations. The differences also did not tend to be systematic. The differences between log and observation measures ranged from 3% to 39% of the absolute differences. This suggests the relatively small discrepancies in the log measures as compared with the observation measures would cancel out even further in larger data sets.

The average difference between log and observation measures of reading time ranged from 1% to 55% as large as the square root of the estimated variance among classes in reading and 3% to 148% as large as the square root of the estimated variance among students within classes in reading. This suggests that one can obtain sufficiently accurate data from teacher logs of reading time for

studying differences among classes and in most cases differences among students within a class. Also the larger the data set the less likely there is to be a problem of measurement error.

In mathematics the average difference between logs and observations ranged from 6% to 26% the average time in mathematics per student per day. The differences between log and observation measures were somewhat more systematic than reading or even language arts. The differences ranged from 21% to 93% of the average absolute value difference.

In all but one class the average difference between the log and observation measures of mathematics time were less than approximately half the size of the square root of the estimated variance among classes in the time spent in mathematics instruction. The estimated variance among students within classes in mathematics instruction was zero. This suggests that teacher logs are in general an acceptable measure of student time in mathematics for studies investigating differences among classes.

In two of the classrooms, the logs and observations differed substantially in the amount of time the recorded that student spending learning science. In one class the observations indicated that students spent on average 14.34 minutes a day in science while the logs indicated they spent no time in science. In the other class, the logs indicated the students averaged 6.02 minutes per day while

the observations indicated they averaged 1.69 minutes per day. In the other four classes the logs and observations were much closer in agreement regarding the time spent in science. The discrepancies in log and observation time ranged from 6% to 29% of the time in science recorded from the observations.

The differences in the time recorded in science from logs and observations was much more systematic than in the other subject matters. In four of the six classes the difference and absolute value difference were equal indicating either the log or observation time was consistently higher than the other across all students on all days included. In the other two classes the ratio of the difference to the absolute difference was 36.

With the exception of one class, the differences between log and observation measures of science time were considerably smaller than the variability among classes in the time per day per student spent on science instruction. The differences ranged from 1% to 26% of the square root of the estimated variance among classes for these five classes. In the sixth class the difference was 86% as large as the estimated variance among classes. Although the differences between logs and observation measures of the time students spent in science may be systematic and not cancel out in large studies, they seem to be small in most cases as compared to the variability among classes in the time spent in science instruction. As in

mathematics, the estimated variance among students within a class in the time spent in science instruction was zero.

Five of the six teachers recorded on average more time in social science than the classroom observers. In two of the classrooms the mean difference between the logs and observations was substantial. In one class there was an average of 5.06 minutes of social studies recorded in the logs per student per day while there was no time in social studies recorded at all in the observations. In the other there was an average of 33.96 minutes recorded in the logs while only 17.88 minutes recorded in the observations. The other four classes had considerably smaller mean differences between the log and observation time in social studies. They ranged from 8% to 37% of the observation time recorded in social studies.

In two classes the average difference between log and observation measures of the time spent in social studies instruction was equal to the absolute value of the average time and only in one class was the difference less than half as large as the absolute value difference. This suggests that like science, the differences in the log and observation measures of social studies time tend to be more systematic than in the other subject matters.

As with science, only in one class (the same class), was the average difference between log and observation time in social studies nearly as large as the square root of the variance among classes. This suggests that teacher

logs probably can in most cases provide a satisfactory source of time on task data for researching differences among classes for social studies. The estimated variance among students within a class was also zero for social studies.

There were substantial differences in the amount of time coded from logs and observations for mixed seatwork in three classes. In one class, over 16 minutes of seatwork on average was recorded from the observations but none was recorded from the logs. In another class almost four times as much time in mixed seatwork was recorded from the logs as from the observations. There also seemed to be a tendency in five of the six classes for the logs or observations to be systematically higher than the other across classes and students. One should note however that in three of the classes the logs indicated more time in mixed seatwork while in the other three the observations indicated more time. Mixed seatwork, as can be seen in the coding procedures presented in appendix B in item 50, constituted a general category for when students were working at their seats on a number of different subject matters and it was not possible to tell upon which subject matters individual students were working. This might explain why there were some large discrepancies in the amount of time coded from logs and observations in this category.

The average difference between the time coded from

logs and observations in transitions between lessons was over 50% as large as the total time coded as transitions between lessons from the observations in three of the six classes. In the three classes with large differences between the amount of time coded from logs and observations, the differences also tended to be systematic, with the average difference per student per day equal to or nearly equal to the average absolute value difference per student per day.

The average difference per day per student coded from logs and observations was greater than the square root of the estimated variance among classes in three of the classrooms, and substantially greater than the variance among students within classes for all six classrooms. This suggests that it may be necessary to use observers to record transition time in classrooms particularly if the researchers are interested in differences among students within a class. It does not seem surprising that teachers seem to have difficulty keeping track of transition time since this is a busy time for them. There also might be some internal pressure for teachers to minimize transition time since it reflects on their classroom management skills, though this might not be the case since three of the six teachers actually recorded more transition time than the observers.

In all but one of the classes more time was recorded in individual activities and less in whole group activities

by the observers as compared to the teachers. The observers also recorded more time on the average than the teachers in subgroup activities for four of the six classrooms. With the exception of the subgroup category in which relatively small amounts of time were coded, the difference between the amount of time coded from logs and observations was small compared to the amount of time coded from the observations.

There were large and consistent differences between the time recorded as teacher supervised and unsupervised activities in the logs and observations for all six classrooms. There seems to be a strong tendency for more time to be recorded as teacher supervised by the teacher as compared to observers. In two classrooms the differences were approximately an hour and a half per day per student. In the other classroom the differences between the logs and observations were also substantial. In five of the classrooms the mean deviation for supervised time exceeded 30 minutes, while in the other classroom it was 17.63 minutes. The differences also were systematic across days and students within a class.

The use of logs does not seem to be an acceptable method of data collecting for the time students spend directly supervised by the teacher. As with transition time, teachers might feel pressure to indicate they are supervising instruction more than they really are. Part of the problem might also be in how the teachers recorded

classroom activities on the logs. An example of the form that was used by the teachers to log student activities is contained in appendix A. As will be discussed in chapter five, it may be possible to improve the ability of teachers to keep track of the time they spend supervising individual students by changing this form.

In summary, a comparison of time allocation data coded from teacher logs and observer notes of the same student activities analyzed at the level of time per day indicates there are more than trivial differences between the two data collection methods in the time they record students spend in different activities. The most drastic difference found was for teacher supervision. There was a consistent tendency across all six classes for teachers to record substantially more student time being directly supervised by themselves than recorded by classroom observers. The differences between log and observation time estimates for the other activity categories, with the exception of reading, were not consistent across classrooms. In reading, although there was a consistent tendency across classrooms for teacher logs on the average to indicate more time in reading, these differences were relatively small.

In reading for all classrooms and for most classrooms in the other categories with the exception of teacher supervision, the absolute mean differences were substantially larger than the mean differences. This indicates

that across students and days, neither logs nor observations recorded more time in a category consistently across days and students. This suggests that while the differences between the logs and observations time recorded in an activity category on a single day for a single student may be large, these differences would tend to cancel out to some extent over days and students in a large scale study.

INDIVIDUAL PURSUIT LEVEL ANALYSIS

When time in different activities measured using teacher logs and observer notes at the level of time per day per student was compared, large differences between the two approaches were found in many cases. In this section the differences between the log and observation measures will be examined at the level of individual pupil pursuits as defined by Harnischfenger and Wiley (1976). Since this is the level at which the activity data is both collected through logs and observations and coded into pursuit records, it should provide better insights as to how differences between log and observation measures of student activities result.

As was described in chapter three, pupil pursuit records were coded from the logs and observations for analysis. They included the beginning and ending time of the pursuit or activity and codings categorizing it in terms of the activity type, student grouping and

supervision. If one or more of these categories changed, the pursuit was considered to have ended and a new one started.

Due to discrepancies between the logs and observations, there was not always a one to one match between the pursuit records coded from logs and observations of a given student's activities on a given day. For example, the log pursuit records might show a student spending from 9:00 to 10:00 studying science individually without teacher supervision, while the observation pursuit records might show the student in transition from 9:00 to 9:05, then studying science individually without supervision from 9:05 to 9:30 then doing seatwork with the whole class under supervision from 9:30 to 10:05. To allow analysis of the discrepancies between log and observation pursuit records, a coding procedure described in detail in chapter three was used. Although the coding procedure was quite straightforward, errors in the coding process would be confounded with true discrepancies between the log and observation pursuit records. For this reason, multiple codings of two student days were done to assess the extent of coder unreliability.

The length of each pursuit record from the two codings was correlated for each of the two student days and found to be .99 and .96 respectively. In addition, the codings for activity type, group type and teacher supervision from the two codings were crosstabulated. For the first

student day that was double coded, of the thirty-four pursuits, there was disagreement for one pursuit each on the three dimensions. For the second student day there was complete agreement for group type, and discrepancies for two pursuits each for activity type and teacher supervision coding for the total of fifty-two pursuits. It is felt that these results suggest that error in coding will have a trivial effect on the results of the analysis at the pursuit level.

PURSUIT LEVEL ANALYSES

As was discussed in chapter three, the hypothesis has been made that there are two major sources of the discrepancies between the logs and observations. First there are differences in the length of pursuit records. A teacher may indicate in his or her log that an activity started at 9:05 and ended at 9:33 while the observer may indicate in his or her notes that the activity started at 9:08 and ended at 9:39. The second major source of discrepancies is due to differences and ambiguities in the descriptions of the student activities contained in the logs and observations resulting in differences in the codes describing the activities in terms of subject matter, grouping and supervision. For example, the same activity may be coded from logs as whole group social studies with teacher supervision while coded as subgroup science without teacher supervision from the observations. Separate

types of analyses were used to assess the extent and nature of these two types of errors. Discrepancies in the coding of the same student activity will be discussed first.

DISCREPANCIES IN PURSUIT CODING

The discrepancies in the coding of pupil pursuits across all classes are presented in Tables 11 through 16. Table 11 presents the coding of the activity type. This includes the major subject matters, transitions between lessons and mixed seatwork. Table 13 includes the coding of the three grouping categories and Table 15 includes the two supervision categories. Each of the tables shows the coding from the logs horizontally and the coding from the observations vertically. The top number indicates the number of pursuits in that cell of the table. The second number indicates the average length of the pursuits in that cell in minutes. The third number indicates the proportion of pursuits with that code from the observations that are in that cell. In other words the row proportions.

A log linear analysis was used to test the hypothesis that the discrepancies between the coding of the three dimensions from logs and observations differed significantly among the teachers. There was a significant lack of fit when the classroom dimension was left out of the model for activity type as well as group and supervision code. The likelihood ratio and Pearson chi-squares and their associated degrees of freedom and probabilities for

the tests of fit on each dimension are given below in Table 10.

Table 10

Differences Among Classes in Pursuit Categorization

	DF	Lr Chisq.	Pearson Chisq.	p <
Activity Type	240	819.43	907.57	.0000
Group Type	40	314.58	346.25	.0000
Teach. Sup.	15	104.85	102.96	.0000

Given the fact that there are statistically significant differences among classrooms in the discrepancies between the logs and observations in how individual pursuit records are coded on the three dimensions, individual crosstabulations by class are provided in Tables 12, 14 and 16. In order to make the tables readable, only the percent of observation pursuits with that coding in the cell are provided. The results contained in Tables 11 and 12 will be discussed first.

The consistency between the coding of activity type was reasonably good for those pursuits coded in the five major subject matter areas. The proportion of pursuits coded from the logs as being in the same subject as coded in the observations ranged from .73 for language arts to .92 for social studies across all classes. There was a tendency to confuse language arts and reading, especially in classes two, three and six. This is not

Table 11

Categorization of Activity Type from Logs and Observations

Observations	Logs						
	Lang. Arts	Read.	Math	Soc. Stud.	Sci.	Tran- sitions	Seat- work
Lang. Arts	244 11.21 .73	35 7.40 .11	3 6.00 .01	0 0.00 .00	0 0.00 .00	7 3.71 .02	44 9.11 .13
Read.	27 11.70 .08	296 11.91 .82	6 22.17 .02	11 16.36 .03	0 0.00 .00	14 5.64 .04	6 18.67 .02
Math	8 5.38 .03	6 8.50 .02	231 11.22 .91	0 0.00 .00	0 0.00 .00	5 2.60 .02	4 28.50 .02
Science	0 0.00 .00	0 0.00 .00	0 0.00 .00	62 14.11 .78	16 38.38 .20	0 0.00 .00	1 11.00 .01
Soc. Stud.	6 24.50 .07	0 0.00 .00	0 0.00 .00	0 0.00 .00	79 11.63 .92	0 0.00 .00	1 3.00 .02
Tran- sitions	90 3.04 .18	46 2.87 .09	39 2.72 .08	32 2.31 .06	27 4.33 .05	243 4.03 .48	27 4.37 .05
Seat- work	49 7.39 .36	27 12.52 .20	1 17.00 .01	0 0.00 .00	1 22.00 .01	20 5.40 .14	40 17.25 .29

Table 12

Categorization of Activity Type from Logs and Observations by Class

Class	Obs.	Logs						
		Lang.	Read.	Math.	Sci.	Soc.	Tran.	Swor.k.
One	Lang.	89	11	0	0	0	0	0
	Read.	7	71	8	0	0	15	0
	Math.	2	2	93	0	0	2	0
	Sci.	0	0	0	100	0	0	0
	Soc.	0	0	0	0	100	0	0
	Tran.	7	2	0	1	0	90	0
	Swor.k.	39	50	0	0	0	11	0
Two	Lang.	70	14	3	0	0	7	7
	Read.	4	96	0	0	0	0	0
	Math.	0	0	100	0	0	0	0
	Sci.	0	0	0	100	0	0	0
	Soc.	0	0	0	0	100	0	0
	Tran.	21	8	12	6	4	47	1
	Swor.k.	48	13	0	0	0	24	15
Three	Lang.	57	24	0	0	0	0	19
	Read.	4	96	0	0	0	0	0
	Math.	2	0	98	0	0	0	0
	Sci.	0	0	0	100	0	0	0
	Soc.	0	0	0	0	96	0	4
	Tran.	15	25	12	19	6	21	2
	Swor.k.	0	33	0	0	0	0	67
Four	Lang.	82	4	0	0	0	4	9
	Read.	0	94	0	3	0	2	0
	Math.	0	8	77	0	0	8	8
	Sci.	0	0	0	100	0	0	0
	Soc.	0	0	0	0	0	0	0
	Tran.	11	16	0	19	0	55	0
	Swor.k.	0	100	0	0	0	0	0
Five	Lang.	95	0	0	0	0	0	5
	Read.	15	35	0	31	0	4	15
	Math.	0	0	100	0	0	0	0
	Sci.	0	0	0	100	0	0	0
	Soc.	32	0	0	0	68	0	0
	Tran.	24	12	12	5	5	24	14
	Swor.k.	16	0	0	0	0	0	84
Six	Lang.	62	3	2	0	0	0	33
	Read.	28	68	0	0	0	0	4
	Math.	11	2	87	0	0	0	0
	Sci.	0	0	0	0	94	0	6
	Soc.	0	0	0	0	100	0	0
	Tran.	26	5	10	0	13	31	15
	Swor.k.	21	0	7	0	7	0	64

surprising given the similarity and overlap of these subject matters. There was also a tendency for pursuits coded as language arts from the observations to be coded as mixed seatwork and visa versa. This was a particular problem in classes two and six. The pursuits coded as math, social studies and science from the observations was for the most part consistently coded the same from the logs. The proportions were .91, .78 and .92 respectively. The major problem in science was in class six where all the 16 pursuits coded as social studies from the observations were coded as social studies from the logs.

Large discrepancies were found for those pursuits coded as transitions from the observations. Only 48% of those pursuits coded from the observations as transitions were also coded as transitions from the logs. The inconsistencies in the coding of transition time occurred in all six classes. This suggests as one might expect, that teachers have trouble keeping track of the time spent in transitions between lessons. In classes one, two and four, there were also a substantial number of pursuits coded from the logs as transitions while being coded as various subject matters and seatwork from the observations. This would explain why there was actually more time coded as transitions between lessons from the logs as from the observations in these classes, as can be seen in Table 8.

Only 29% of the pursuits coded as seatwork from the

observations were also coded as seatwork from the logs. These inconsistencies occurred across all the classrooms though it was a particular problem in classes one, two, and four. In general, seatwork was confused with reading and language arts. There was also a strong tendency for pursuits coded as subject matters and transitions from the observations to be coded as seatwork from the logs. Of those pursuits coded as seatwork from the logs, 67% were coded as other than mixed seatwork from the observations. As was stated earlier, the mixed seatwork category was used when students were working on multiple subject matters at their seats and it was not possible to determine which subject matter a particular student was working on at any given time. It is not surprising that there were a large number of pursuits where there was disagreement in terms of this category.

The crosstabulation of group type coded from logs and observations at the pursuit level is presented in Tables 13 and 14. The major discrepancy was in the subgroup category. Of those pursuits coded as subgroup from observations, over half were coded as individual pursuits in classes three through six from the logs. In class one, six of the eight pursuits coded as subgroup from the observations were coded as whole group from the logs. The whole group and individual categories were found to be more consistently coded from logs and observations. Approximately 80% of those pursuits coded as whole group

Table 13

Categorization of Group Type from Logs and Observations

Observations	logs		
	Whole Group	Subgroup	Individual
Whole Group	385 15.63 .79	10 9.80 .02	95 6.88 .19
Sub-Group	8 25.13 .07	42 13.88 .39	59 10.58 .54
Individual	88 11.48 .12	45 5.69 .07	575 12.67 .81

Table 14

Categorization of Group from Logs and Observations by Class

Class	Obs.	Log		
		Whole	Sub.	Ind.
One	Whole	68	0	32
	Sub.	75	12	12
	Ind.	7	0	93
Two	Whole	88	4	8
	Sub.	0	76	24
	Ind.	16	8	79
Three	Whole	83	0	17
	Sub.	6	6	88
	Ind.	18	0	82
Four	Whole	98	0	2
	Sub.	5	43	52
	Ind.	14	29	57
Five	Whole	53	7	40
	Sub.	0	34	66
	Ind.	15	0	85
Six	Whole	70	1	29
	Sub.	0	44	56
	Ind.	4	2	94

and individual activities from observations were coded the same way from the logs.

The crosstabulation of teacher supervision coded from logs and observations at the pursuit level is presented in Tables 15 and 16. There was a strong tendency for those pursuits coded as not being supervised by the teacher from the observations to be coded as supervised from the logs. In all six classes over half the pursuits coded from the observations as unsupervised were coded as supervised from the logs. Among those pursuits coded as supervised from the observations, 88% were coded as supervised from the logs. Given these results at the pursuit level, it is not surprising that there was substantially more student activity time categorized as teacher supervised in all the classrooms when the data was aggregated to the level of student time per day.

The discrepancies in the categorization of the pursuits from logs and observations at the pursuit level seems to parallel quite closely the discrepancies in the average time in those categories coded from logs and observations at the level of total time per student day discussed previously in this paper. This suggests that miscategorization of pupil pursuits is a major source of the discrepancies between the time students spend in different activities measured using teacher logs and observer notes. The other potential source of discrepancies between the time students spend in different activities

Table 15

Categorization of Supervision from Logs and Observations

Observation	Log	
	Supervised	Unsupervised
Supervised	715	98
	11.90	5.33
	.88	.12
Unsupervised	302	152
	15.12	15.38
	.67	.33

Table 16

Categorization of Supervision from Logs and Observations

Class	Obs.	Log	
		Sup.	Unsup.
One	Sup.	90	10
	Unsup.	70	30
Two	Sup.	77	23
	Unsup.	52	48
Three	Sup.	97	3
	Unsup.	80	20
Four	Sup.	84	16
	Unsup.	68	32
Five	Sup.	96	4
	Unsup.	81	19
Six	Sup.	88	12
	Unsup.	63	37

measured using logs and observations is differences in the recorded time pursuits start and end. The analysis of this source of error is discussed next.

DISCREPANCIES IN THE LENGTH OF PURSUITS

A measurement model discussed by Schmidt (1981) was used to assess how consistent teachers were with observers in recording the beginning and ending times of student activities. The model which was described in chapter three is essentially the linear regression of a observed score on its true score. In this instance, the pursuit length computed from the observer recorded beginning and ending times is defined as the true score and the pursuit length computed from the teacher recorded beginning and ending times is defined as the observed score. Schmidt shows algebraically that error or the difference between the true and observed score contains three distinct components. One is a fixed bias independent of the true score and equal to the intercept of the observed score regressed on the true score. The second is a random bias associated with the magnitude of the true score and equal to the slope of the regression of the observed score on the true score times the true score minus one. The third component is random error independent of the true score and equal to the standard error of estimate of the regression of the observed score on the true score. The pursuit lengths from both the observations and the logs (true and observed

scores in the model) are random variables. For this reason, maximum likelihood was used to estimate the parameters and provided correct asymptotic standard errors.

The model was fit separately for the pursuits in each class. It was felt that teachers were likely to differ in how accurately they could keep track of the beginning and ending times of the activities of the students in their classes. These results are presented in Table 17. The table contains the mean pursuit length in minutes from the logs and observations, fixed bias, the random bias coefficient (one minus the slope of the log time regressed on the observation time), square root of the random error and the number of pursuits. The table also contains the asymptotic standard errors for the fixed bias component and random bias coefficient. The model was also fit separately for the pursuits in the five subject matters, transitions between lessons, seatwork, group type and whether or not the pursuit was supervised by the teacher. Only those pursuits that were coded the same from logs and observations were included. These results are presented in Table 20 which contains the same statistics as Table 17. In order to give some perspective to the effect of fixed and random bias when using teachers as opposed to observers for recording the beginning and ending times of pursuits, the fixed, random and total bias for pursuit lengths of 15, 30, 45 and 60 minutes for each class and activity category are presented in Table 19. Fixed, random and

Table 17

Sources of Error in Pursuit Length Coded from Teacher Logs by Class

Class	Mean	Time	Obs. Log Bias	Fixed Coefficient	Random Bias Stand. Dev.	Error Count
1	10.78	10.29	0.38 (.20)	-.08 (.01)	2.91	399
2	7.02	6.94	0.05 (.10)	-.02 (.01)	2.02	616
3	10.96	10.65	0.33 (.25)	-.06 (.01)	4.21	445
4	14.40	13.32	1.75 (.38)	-.20 (.02)	5.43	350
5	14.24	14.05	0.54 (.28)	-.05 (.01)	3.12	251
6	11.87	11.56	-0.55 (.22)	.02 (.01)	3.80	510

total bias at given pursuit lengths are also provided for each of the activity categories in Table 21. Shorter sample pursuit lengths (1, 2, 5 and 10 minutes) are used for transitions since, as can be seen by the mean log and observation times presented in Table 20, these pursuits tend to be shorter.

Error of measurement or in this case the difference between teacher and observer measures of pursuit length is defined under the model in equation 1

$$e = A + (B-1)\xi + \epsilon \quad (1)$$

where e is error, A is fixed bias, B is random bias, and ϵ equals random error. Since A is a fixed constant, the variance of e is given below in equation 2.

$$\sigma_e^2 = (B-1)^2 \sigma_\xi^2 + \sigma_\epsilon^2 \quad (2)$$

Table 18 presents the total error variance, random bias variance and random error variance for each of the six classrooms. The proportion of the error variance from random bias and error are also presented. The same statistics are also presented in Table 22 for each of the activity categories.

The results of fitting the model across all types of pursuits in each classroom will be discussed first.

The fixed bias among the teachers tended to be small and with the exception of classroom six, positive. Only classroom four had a fixed bias exceeding one minute. Only

Table 18

Bias at Standard Pursuit Lengths by Class

Class	Pursuit Minutes	Total Bias	Fixed Bias	Random Bias
1	15	-.82	.38	-1.20
	30	-2.02	.38	-2.40
	45	-3.22	.38	-3.60
	60	-4.42	.38	-4.80
2	15	-.25	.05	-.30
	30	-.55	.05	-.60
	45	-.85	.05	-.90
	60	-1.15	.05	-1.20
3	15	-.57	.33	-.90
	30	-1.47	.33	-1.80
	45	-2.37	.33	-2.70
	60	-3.27	.33	-3.60
4	15	-1.25	1.75	-3.00
	30	-4.25	1.75	-6.00
	45	-7.25	1.75	-9.00
	60	-10.25	1.75	-12.00
5	15	-.21	.54	-.75
	30	-.96	.54	-1.50
	45	-1.71	.54	-2.25
	60	-2.46	.54	-3.00
6	15	-.25	-.55	.30
	30	.05	-.55	.60
	45	.35	-.55	.90
	60	.65	-.55	1.20

in classrooms four and six did the fixed bias differ from zero by more than two standard errors. The random bias coefficient among the teachers was also small and with the exception of classroom six, negative. The magnitude of the random bias coefficient was under .10 with the exception of classroom four where it was .20. The random bias coefficients though small, differed from zero by more than two standard errors in all but classrooms two and six. The standard deviation of the random error which can be thought of as roughly the average amount of random error in a teacher recorded pursuit, ranged from 2.02 minutes in classroom two to 5.43 minutes in classroom four.

As stated above, with the exception of classroom six, the fixed bias was positive and the random bias negative. In classroom six the signs of the bias coefficients were reversed. This suggests that in the first five classes the teachers tend to underestimate short pursuits and overestimate long pursuits. In classroom six the teacher tended to underestimate short pursuits and overestimate long pursuits. This is due to the random bias being a function of the true or observation pursuit length, the longer the pursuit the greater the magnitude of random bias. The fixed bias is constant across pursuit of different lengths. This would result in the magnitude of the fixed bias to be greater than the magnitude of the random bias up to a given pursuit length, while the magnitude of the random bias would be greater for larger

pursuits. Through simple algebra it can be shown that the magnitude of the fixed bias is equal to the magnitude of the random bias when the pursuit length is equal to $A/(B-1)$, where A is the fixed bias and B-1 is the random bias coefficient. At this pursuit length, if the fixed and random bias differ in sign they cancel out resulting in no bias. If the fixed and random bias were of the same sign, they of course would be additive.

Table 19 presents the total, fixed and random bias for 15, 30, 45 and 60 minute pursuits in each classroom. The total bias is negative across all four pursuit lengths in the first five classrooms and is negative for 15 minute pursuits and positive for larger pursuits in classroom six. The bias for a 60 minute pursuit was -10.25 minutes in classroom four. In the other five classrooms the magnitude of the bias for a 60 minute pursuit was under five minutes. Only in classroom four did bias seem to be a major factor in teacher recorded pursuit length. This is also the only classroom where both the fixed bias component and the random bias coefficient differed from zero by more than two standard errors.

Table 19 presents the random bias random error and total measurement error variances for each classroom. These results indicate that random errors in recording pursuit length as opposed to random bias seems to be the major contributing factor to measurement error variance. Only in classroom four did the variance from random bias

Table 19

Bias and Error Components of Measurement Error Variance by Classroom

Class	Measurement Error Variance	Random Bias	Random Error
1	9.35	.90 (10)	8.45 (90)
2	4.11	.03 (1)	4.08 (99)
3	18.44	.72 (4)	17.72 (96)
4	40.17	10.69 (27)	29.48 (73)
5	10.25	.52 (5)	9.73 (95)
6	14.51	.07 (1)	14.51 (99)

account for more than 10% of the total error variance, and in that class it accounted for under a third of the variance.

Table 20 presents the results of fitting the model for different types of pursuits. This analysis was collapsed across classes to ensure reasonable sample sizes and keep the sets of results to a manageable number. As when the model was fit across all types of pursuits, bias did not seem to be a significant factor for recording pursuit length in the five subject matters. The fixed bias was positive and under a minute in all five areas. Only in math and language arts did it differ from zero by more than two standard errors. The random bias coefficient was negative and less than .1 for all the five subjects. Only in reading and math did it differ from zero by greater than two standard errors. These results suggest that for science and social studies, the classical true score model may be appropriate for modeling error in teacher recorded pursuit length. Since under the classical true score model error is random with an expected value of zero, it would tend to cancel out in large scale studies. This would tend to indicate that for those pursuit categories where the classical true score model fits, the error in teacher recorded pursuit length should not be a problem.

Table 21 presents the fixed, random and total bias estimates for the five subject matters as well as

Table 20

Sources of Error in Pursuit Length Coded from Teacher Logs by Activity

Class	Obs. Log Mean Time		Fixed Bias	Random Bias Coefficient	Error Stand. Dev.	Count
Lang. Arts	11.21	11.51	.52 (.19)	-.02 (.01)	2.08	244
Reading	11.91	11.82	.29 (.21)	-.03 (.01)	2.75	296
Math	11.17	11.26	.88 (.30)	-.07 (.02)	3.50	230
Science	14.11	14.40	.32 (.96)	-.01 (.05)	5.41	62
Soc. Stud.	11.63	12.17	.62 (.53)	-.01 (.03)	3.49	79
Transitions	4.03	4.49	1.75 (.23)	-.32 (.05)	1.91	241
Seatwork	17.25	16.87	2.77 (.82)	-.18 (.04)	4.19	38
Whole Group	15.63	15.79	.92 (.26)	-.05 (.01)	3.59	385
Subgroup	13.88	14.09	-.18 (.98)	.03 (.05)	4.28	42
Individual	12.67	12.68	.75 (.21)	-.06 (.01)	3.51	573
Supervised	11.90	11.97	.59 (.15)	-.04 (.01)	2.95	715
Unsupervised	15.38	15.01	2.08 (.69)	-.16 (.03)	5.40	152

Table 21
Bias at Standard Pursuit Lengths by Activity

Activity	Pursuit Minutes	Total Bias	Fixed Bias	Random Bias
Language Arts	15	.22	.52	-.30
	30	-.08	.52	-.60
	45	-.38	.52	-.90
	60	-.68	.52	-1.20
Reading	15	-.16	.29	-.45
	30	-.61	.29	-.90
	45	-1.06	.29	-1.35
	60	-1.51	.29	-1.80
Math	15	-.17	.88	-1.05
	30	-1.13	.88	-2.01
	45	-2.27	.88	-3.15
	60	-3.32	.88	-4.20
Science	15	.17	.32	-.15
	30	.02	.32	-.30
	45	-.13	.32	-.45
	60	-.28	.32	-.60
Social Studies	15	.47	.62	-.15
	30	.32	.62	-.30
	45	.17	.62	-.45
	60	.02	.62	-.60
Transitions	1	1.43	1.75	-.32
	2	1.11	1.75	-.64
	5	.15	1.75	-1.60
	10	-1.45	1.75	-3.20
Mixed Seatwork	15	.07	2.77	-2.70
	30	-2.63	2.77	-5.40
	45	-5.33	2.77	-8.10
	60	-8.03	2.77	-10.80

Table 21 (Cont'd.)

Activity	Pursuit Minutes	Total Bias	Fixed Bias	Random Bias
Whole Group	15	.17	.92	-.75
	30	-.58	.92	-1.50
	45	-1.33	.92	-2.25
	60	-2.08	.92	-3.00
Subgroup	15	.27	-.18	.45
	30	.72	-.18	.90
	45	1.17	-.18	1.35
	60	1.62	-.18	1.80
Individual	15	-.15	.75	-.90
	30	-1.05	.75	-1.80
	45	-1.95	.75	-2.70
	60	-2.85	.75	-3.60
Supervised	15	-.01	.59	-.60
	30	-.61	.59	-1.20
	45	-1.21	.59	-1.80
	60	-1.81	.59	-2.40
Unsupervised	15	.32	2.08	-2.40
	30	-2.72	2.08	-4.80
	45	-5.12	2.08	-7.20
	60	-7.52	2.08	-9.60

transitions and mixed seatwork. Of the five subject matters, only in reading and math does the total bias exceed a magnitude of one minute for even a 60 minute pursuit. For mathematics and reading the magnitude of the bias was under five minutes for a 60 minute pursuit. This suggests that even in the two subject matters where bias differed from zero by more than two standard errors, the magnitude of the bias is relatively slight.

Table 22 presents the random bias and random error components of total measurement error or log time minus observation time variance for the 12 activity categories. In science and social studies, random bias accounted for less than one percent of the total measurement error. In the other subject matter categories random bias also accounted for a very small portion of the total error variance. This coupled with the results presented above, suggests that the major source of error in teacher recorded pursuit length is random error rather than bias.

Bias was a more important factor in recording pursuit length for transitions between lessons. The fixed bias was 1.75 minutes and the random bias coefficient was $-.32$. The fact that the fixed bias and the random bias coefficient differed in sign indicates they would tend to cancel each other out with no bias when the pursuit length was 5.46 minutes. Both the fixed bias and random bias coefficient differed from zero by well over two standard errors. These results suggest that teachers have difficulty

Table 22

Bias and Error Components of Measurement Error Variance by Activity

Activity	Measurement Error Variance	Random Bias	Random Error
Language Arts	4.39	.05 (1)	4.33 (99)
Reading	7.73	.17 (2)	7.56 (98)
Mathematics	13.08	.83 (6)	12.25 (94)
Science	29.29	.02 (0)	29.27(100)
Social Studies	12.20	.02 (0)	12.18(100)
Transitions	4.27	.62 (15)	3.65 (85)
Seatwork	27.96	10.40 (37)	17.56 (63)
Whole Group	13.47	.58 (4)	12.89 (96)
Subgroup	18.47	.15 (1)	18.32 (99)
Individual	12.91	.58 (4)	12.32 (96)
Supervised	8.99	.29 (3)	8.70 (97)
Unsupervised	33.54	4.38 (13)	29.16 (87)

keeping track of the length of time their students spend in transitions. Teachers tend to overestimate short transitions (less than 5.46 minutes) and underestimate longer pursuits. This is not surprising given that teachers are generally quite busy managing the classroom during these periods and probably do not have time to look at the clock and record beginning and ending times. Table 21 presents the estimated fixed, random and total bias for transition pursuits of 1, 2, 5 and 10 minute lengths. Shorter pursuit lengths were given for transitions in Table 21 due to the fact that transitions tend to be shorter than other activities as can be seen from the mean observation and log pursuit lengths given in Table 20. As one can see the estimated bias in teacher recorded pursuit length for a one minute pursuit is 1.43 minutes, greater than the actual pursuit length. For a 10 minute pursuit, the total bias has reversed sign and is -1.45 minutes. Random bias accounted for 15% of the error variance for transitions between lessons. This was far greater than was found for the five subject matter areas, though random error still seems to be the major contributor to error variance in teacher recorded pursuit length.

Bias was also a factor in the teacher's recording of the pursuit length of seatwork. The fixed bias was 2.77 minutes and the random bias weight was -.18. Both differed from zero by well over two standard errors. Random bias also made up a significant portion of the total error

variance. Apparently teachers are more biased in keeping track of the time their students spend doing mixed seatwork than they are in keeping track of the time their students spend in the five subject matter areas. As can be seen in Table 22, teachers tend to slightly overestimate short mixed seatwork pursuits (.07 minutes for a 15 minute pursuit) and underestimate longer pursuits (8.03 minutes for a 60 minute pursuit).

There was only a small amount of fixed and random bias in the recording of the pursuit length for all three grouping categories. Both the fixed bias and random bias weight for the subgroup category differed from zero by less than one standard error. The fixed bias for whole group and .75 individual pursuits was .92 and respectively. Both were larger than three standard errors. The random bias weights were -.05 and -.06 for whole group and individual pursuits respectively, both differing from zero by more than three standard errors. Despite the fact one can be quite confident that random bias differed from zero for whole group and individual activities, it made up only a small portion of the total error variance. As can be seen in Table 22, random bias made up only 4% of the error variance for both individual pursuits and whole group pursuits. As can be seen in Table 21, the total bias for a 60 minute pursuit was less than three minutes for all the grouping categories. The square root of the random error variance for whole group, subgroup and

individual pursuits was 3.59, 4.28 and 3.51 minutes respectively. As with the five subject matters, this was about 20% to 30% as large as the average pursuit length.

As one might expect, teachers had more difficulty keeping track of the beginning and ending times of unsupervised pursuits as compared with supervised pursuits. There was both more bias and random error for unsupervised pursuits. The fixed bias for supervised and unsupervised pursuits was .59 and 2.08 minutes respectively and in both cases the fixed bias estimates were larger than three standard errors. The random error coefficient for supervised and unsupervised pursuits were -.04 and -.16 respectively and both differed from zero by more than three standard errors. The standard deviation of the random error variance was also larger for unsupervised as opposed to supervised pursuits, 5.40 to 2.95 minutes respectively. For both categories, random error variance as opposed to random bias tended to make up the major portion of the total measurement error variance.

When pursuit data was aggregated to the level of total time per day per activity category, the teacher logs indicated considerably more time spent as supervised and less as unsupervised than the observer field notes. As was discussed in the previous section, this was at least partly explained by differences in the categorization of pursuits. The results in Table 20 suggests it is also partly explained by a tendency for teachers to overestimate

on the average the length of supervised pursuits and underestimate the length of unsupervised pursuits. The mean pursuit length of a supervised pursuit from the observations is slightly smaller than from the logs (11.90 to 11.97). The reverse is true for unsupervised pursuits to a greater extent (15.38 to 15.01). It is difficult to evaluate the impact of this tendency or bias at the level of time per day supervised and unsupervised, since there may be many supervised and unsupervised pursuits per student per day. What seems clear is that the tendency for teachers to underestimate the length of unsupervised pursuits and overestimate the length of supervised pursuits to some extent along with miscategorization, results in considerably more student time being recorded as teacher supervised from teacher logs than from observer notes.

In summary, a model with a fixed and/or random bias component is necessary for some teachers and some activity or pursuit categories across teachers, while in others the classical true score model is adequate. Even in those categories, or for those teachers where the random bias was statistically significant, random error made up the major portion of the total error variance. The random and fixed bias point estimates differed in sign for each teacher and each pursuit category. With the exception of teacher four and subgroup instruction, the fixed bias was positive and the random bias negative. This indicates

that in most cases teachers tend to underestimate longer pursuits and overestimate shorter ones.

The square root of the random error variance ranged from about 20% to 30% as large as the average pursuit length across teachers and pursuit categories. This component can be thought of as roughly the average difference between teacher and observer recorded pursuit length due to random errors. Though substantial, this component which is random with an expected value of zero under either Schmidt's model or the classical true score model, would tend to cancel out when averaged over pursuits.

CHAPTER FIVE

CONCLUSIONS AND IMPLICATIONS FOR MEASURING TIME ON TASK

This chapter begins with a summary of the results of assessing the use of teachers to collect classroom time allocation data. The results will be discussed in the order they were presented in chapter four. This will be followed by a discussion of the implications of the results for future studies of time allocation and how data collection procedures might be improved.

SUMMARY OF THE RESULTS

For four days each in two of the classrooms two individuals coded the observer transcripts in order to assess the reliability of the coding procedures. The time in five subject matter areas coded by two different coders from the same observer transcripts was fairly consistent. In most cases, the average difference between coders as well as the average of the absolute value of the difference was less than 10 percent of the time coded for that subject matter. The coders had some difficulty coding mixed seatwork, suggesting there may be some confusion as to what is mixed seatwork and what is instruction in a given subject matter. One set of coders had large

differences in the amount of time they coded in transitions between lessons. The other coders were quite consistent in the time coded as transitions. There were large differences among some of the pairs of coders in the time students spent in the different grouping categories. The most substantial difference was between one pair of coders for the subgroup category where one coded approximately 24 minutes a day and the other coder coded essentially no time in this category. The fact the other three pairs of coders were fairly consistent in the time they coded in this category suggests this was probably an isolated problem. All 4 pairs of coders were fairly consistent in the time coded in the supervisory categories. There also did not seem to be a consistent tendency of one coder or the other to record more time as either supervised or unsupervised across days and students.

The coders in general were quite consistent in the time they coded in the different activity categories. Although there were cases in which a pair of coders differed substantially in the time they coded for a specific category, there were no categories where a majority of the four pairs of coders differed substantially in the time they coded in the category. This suggests that differences between coders in the time they code is probably idiosyncratic, possibly having more to do with differences in how they coded a specific activity, rather than any general tendency. Only two days of a single

class were coded by each pair of coders due to the substantial amount of time and effort it took to code observer notes. If a single activity that a large portion of the class participated in was coded differently by the coders, this single discrepancy could result in substantial differences in the time coded by the coders in the two categories.

Generalizability theory was used to estimate the reliability of a number of sample data collection designs. This provided insights on how best to collect reliable data in a cost effective fashion. The facets were measures (observations and logs), classes, days and students. The time categories were language arts, reading, math, science, social studies and transitions between lessons. Estimates of the variance components were computed for each facet and all the existing interactions among the facets. Classes, days and the measure by day interaction tended to be the largest components. The major exception was the class by measure interaction which was the largest component for transitions.

The reliability for measuring differences among classes was estimated for all possible combinations of using logs or observations, two or four days, and five or ten students. The reliability for measuring differences among students within a class was estimated for all combinations of logs or observations and two or four days.

The results suggest that increasing the number of days

observed is generally the most effective method of increasing reliability, though using observers as opposed to teachers results in a substantial improvement. Increasing the number of students observed had little or no effect on the reliability. It seems clear that reasonably reliable data, e.g. coefficients of .80, can be obtained with the use of either teachers or observers as the data source and a reasonable number of days observed, e.g. under fifteen, if the focus is on assessing differences among classes. If, however, a research is interested in measuring differences among students within classes, the task is much more difficult if not impossible. There was no variance among students for science and social studies and hence differences among students cannot be measured. In the other categories there were relatively small amounts of variance among students. Only in reading and possibly language arts could differences among students be measured reliably and only through the use of observers on a large number of days.

For eight days in three of the classrooms and nine days in the other three classrooms observers as well as the classroom teachers recorded student activities and the time they occurred. The amount of time coded from the teacher logs and observer transcripts was compared for five subject matters, transitions between lessons, seat-work, group type and whether or not the activity was supervised by the teacher. Although the differences between

the time coded from the logs and observations probably contain differences due to coder error as well as differences in the information recorded by the teacher and observer, the differences found in the time in different categories between the log and observation data far exceeded that between multiple codings of the same observation discussed above. This indicates that there are real and substantial differences between the information collected by teachers and observers.

The most striking and consistent difference across all the classrooms was the tendency for teachers to record more supervised time and less unsupervised time as compared with the observers. This was true for all teachers. The difference ranged from 11.51 minutes per day less time as unsupervised being recorded by one teacher to a high of 104.48 minutes per day more time being recorded as supervised by another teacher. If the assumption is made that the error is in the teacher logs rather than the observer notes, these results indicate the use of teachers to record the time their students spend under their direct supervision will not provide accurate data.

Reading was the only other category where there were consistent differences between the logs and observations. In all the classrooms the teachers recorded more time in reading than the observers, though the differences were nowhere near as great as in the supervision categories. The fact the mean absolute value of the difference between

logs and observations was substantially larger than the mean difference in all the classrooms indicates that none of the teachers consistently recorded more time than the observers across all the days and students in his/her class in reading.

In all the other activity categories as well as the three group type categories, there was no consistent tendency across classes for more time to be recorded by the teachers or the observers. The magnitude of the discrepancies between log and observation time varied widely from classroom to classroom. Also there was no classroom where the discrepancies were consistently large or small across the categories as compared with the other classrooms, suggesting that there were no teachers that were particularly accurate or inaccurate in recording student activities.

In order to better understand the differences that existed between the teacher logs and observer notes, the discrepancies were also analyzed at the level of individual pursuit records. Given the nature of the pursuit records, a coding process was used to create a one to one match between the log and observation pursuits. Given the size of the data set, this was done only on a random sample of days in each classroom for a random sample of students within each group of students that had similar activities. Multiple codings of two student days were done to assess the reliability of this coding process and it was found to be quite reliable.

Two types of errors could exist between pursuit records coded from the logs and observations. First they could differ as to how the pursuit was categorized in terms of the type of activity and/or the group and supervision type. Secondly, differences could exist in the recorded length of the pursuit. Separate procedures were used to assess each type of difference. A crosstabulation of the codings of each pursuit from the logs and observations was done both across classes and within classes for activity type as well as group and supervision type. A measurement model developed by Schmidt (1981) was used to assess the discrepancies in pursuit length recorded in the observations and logs.

The consistency between the log and observation data was quite good for the categorization of the five subject matter areas. This was not the case for transitions between lessons and seatwork. The log and observation codings were reasonably consistent for whole group and individual instruction. This was not the case for subgroup activities where the majority of pursuits coded in this category from the observations were coded as individual pursuits from the logs. As one might expect from the results at the level of total time per day per student, the majority of pursuits coded as unsupervised from the observations were coded as supervised from the logs.

Schmidt's model allowed the error or discrepancies in pursuit length between the logs and observations to be

partitioned into three distinct components. The first is a fixed bias in estimating pursuit length from the logs as compared with the observations. The second is a random bias associated with pursuit length. The third is a random error independent of pursuit length. The model was fit for each class across all pursuits and for individual pursuit categories across all teachers.

With the exception of classroom six, there was a slight positive fixed bias and a slight negative random bias. In classroom six the signs were reversed. This indicates that in the first five classrooms the teachers tended to overestimate short pursuits and underestimate long pursuits. In classroom six this tendency was reversed. With the exception of classroom four, these tendencies were small. There was a substantial amount of random error in estimating pursuit length in all six classrooms. The standard deviation of the error variance ranged from about 20% to 30% of the average length of a pursuit or from about 2 to 5 minutes across the classes.

As when the model was fit for the pursuits in each class, there was in general a slight positive fixed bias and slight negative random bias when the model was fit for different types of pursuits with the following exceptions. Both the random and fixed bias component were substantially higher for transitions between lessons, seatwork and unsupervised activities. This suggests teachers have greater difficulty keeping track of these types of student

activities, tending to overestimate the length of the short pursuits and underestimate the long pursuits. The sign of the bias components was reversed for subgroup activities though the magnitude of both components was small, indicating a slight tendency for teachers to underestimate the length of short pursuits and overestimate the length of long pursuits for subgroup activities. There was a substantial amount of random error in estimating pursuit length for all the types of activities. The amount of random error ranged from about 20 percent of the average pursuit length in reading to almost 50 percent of the average pursuit length in unsupervised activities.

CONCLUSIONS

The major purpose of this study was to evaluate teachers as a source of classroom time allocation information. The first question this study addressed was how reliable was the coding procedure used by the Language Arts Project to transform written descriptions of student activities into pursuit records, that is categorizing the activities and recording their length. The second question was what were the major sources of variance in time allocation and how do they affect the reliability of various data collection designs. The third and major question of this study was to what extent can teachers provide accurate information on their student's activities by keeping daily logs.

CODER RELIABILITY

As stated above, pursuit records from two codings of the same observer transcripts were found to be in general fairly consistent though there were some exceptions. In addition, in most cases one of the coders in each pair was not consistently higher than the other across students and days coded. This would suggest that the differences would to some extent cancel out when coding large data sets. Given these results, it is probably not necessary to use multiple coders in the future on a regular basis using the procedures developed by the Language Arts Project. This could potentially save a great deal of effort given that it takes about eight hours to code a class day if all the students in the class are observed. Since there were categories where some of the pairs of coders did disagree substantially, it would be wise to double code a few class days coded by each coder to locate problems if they exist. It would also be a good idea to rotate coders so that a class was coded by more than one individual. This would help balance out individual coder biases. The consistency of the multiple codings also suggests the pursuit records coded from the descriptions of classroom activities collected by the Language Arts Project are reasonably accurate and can provide useful information on how student time is allocated in elementary classrooms.

THE GENERALIZABILITY OF MEASURES OF TIME ALLOCATION

The major sources of variance in measures of time allocation found in this study were classes, days and the interaction of measures and days. The one exception was for transition time where the largest source of variation was the class by measure interaction. The implications of generalizability study results in terms of the accuracy of teacher logs will be discussed in the next section.

When the variance components were used to compute the reliability of a number of sample research designs, the results made it clear that both increasing the number of days observed and using observers as opposed to teachers resulted in substantial increases in reliability. Increasing the number of students observed within a class has little or no effect on the reliability. It also became clear that researchers should have no trouble obtaining reliable data when they are interested in comparing classrooms in terms of the time their students spend on average in various activities.

There was little, and in the case of science and social studies, no variance among students within a class. Obviously it is not possible nor necessary to measure differences in time allocation among students when the differences do not exist as the results of this study suggest for science and social studies instruction. There was however variation among students in language arts, reading, math and transitions. The reliability

coefficients from the sample data collection designs suggest that it would be possible to obtain reasonably reliable data for comparing time differences in reading and possibly language arts among students within a class. It would however require sampling a large number of school days, particularly if teachers as opposed to observers were used as the data source. This was not the case for math and transitions. There was so little variability among students in a class for these activities that it would require collecting data on a huge number of days to obtain reliable data even if observers were used. The differences among students in a class in the time spent studying math or in transition between lessons is so small however that it is probably not worth investigating anyway.

THE ACCURACY OF TEACHER LOGS

As was discussed above, it seems clear that teacher logs can provide time allocation data accurate enough to measure differences among classes in broad subject matter categories. Although there was roughly a .10 improvement in reliability when observers as opposed to teachers were used as the data source, this difference could be made up by increasing the number of days data is collected. This may well turn out to be a more cost effective approach.

Research on teaching often focuses on much more narrowly defined student activities than subject matter

categories such as math or reading. The researchers of the BTES study (Fisher, et al., 1978) for example, measured time allocation in many categories within math and reading such as word structure and computational transfer. The Harnischfenger and Wiley model defines pupil pursuits in terms of the intersection of subject matter grouping and supervision. It is not clear from this study whether or not teachers can provide accurate data when the time categories are more narrowly defined. Teachers grossly overestimated the amount of time they spend directly supervising student instruction. There was also error in time they recorded students spent in different grouping categories. When student activities or pursuits are defined in terms of the intersection of subject matter group and supervision, it is questionable as to whether teachers are of use. This study of course did not address the accuracy of teacher logs for recording student time in narrowly defined subject matter categories.

In addition to the question of whether teacher logs are accurate enough to be used for collecting time allocation data, this study provided insights about the nature of the error in teacher logs. The analysis at the pursuit level found the discrepancies between the observers and teachers were due to both differences in how individual pursuits were categorized as well as their length. It is difficult to assess the impact of each of the two types of error, though it seems likely that the errors in

categorization of pursuits had a greater effect for two reasons. First, the discrepancies in categorization that were found to a large extent seemed to explain the discrepancies between the log and observation time in many of the categories found in the analysis at the level of time per student per day. This was true even though the pursuit level analysis was done on only a small sample of the data as the analysis at the level of time per student per day. The most striking example of this is supervision time where there was a tendency for pursuits categorized as unsupervised from observations to be categorized as supervised from the logs. The extent this occurred in each classroom is roughly related to the size of the discrepancy found when the pursuits were aggregated to the time per student per day. Secondly, the results of fitting Schmidt's model suggests that there was little fixed or random bias in most of the pursuit categories. Although there was a substantial amount of random error in pursuit length, these differences being random would tend to cancel out.

There is probably little that can be done to improve teacher accuracy in recording the beginning and ending time of pursuits. Their prime responsibility is teaching and during busy periods they are likely to have to guess the exact time a pursuit started or ended. As stated above, the analysis at the pursuit level suggests that the discrepancies between the teachers and observers in recording beginning and ending times of the pursuits is

mainly random and would tend to cancel out. This suggests that although there is little that can be done to improve the accuracy of teachers in recording the times pursuits begin and end, nothing really needs to be done.

Differences in the categorization of pursuits seemed to be the major source of discrepancies in the time recorded in different categories from log and observation descriptions. As can be seen in the examples of teacher logs and observer transcripts presented in appendix A, the descriptions of the pupil's activities in the observer transcripts provide far greater detail than the descriptions in the teacher logs. It seems reasonable to assume that the discrepancies in the categorization of pursuits is due mainly to errors in coding of the logs because the coders were not provided with enough information. Although it is unreasonable to expect teachers to provide the kind of detail contained in the observations, changes in the log forms the teachers use could possibly improve the accuracy of the information they provide.

The greatest discrepancies in pursuit categorization between the logs and observations were in teacher supervision. There was also a strong tendency for pursuits coded as subgroup from observer transcripts to be coded as individual from teacher logs in four of the six classes. One possible way of improving the accuracy of categorizing the grouping and supervision dimensions is to have the teacher code them directly on the logs. Columns could be

provided for each dimension into which the teacher could put a code indicating the category; i.e. 1 for supervised or 2 for unsupervised. This procedure might also be feasible for the activity type (subject matters, transitions, etc.) though the number of categories is greater and the distinctions between them are less clear. Direct coding of pupil pursuits by the teachers on some or all the dimensions the pursuits are categorized would also reduce the work or eliminate the need for a separate coding process. Field testing of this approach for teacher log keeping would be necessary to see if it is effective for categorizing along each specific dimension. Using teachers to directly code pursuits would also have the advantage of reducing the work or actually eliminating the need for the coding process.

APPENDICES

APPENDIX A

APPENDIX A

Observation

- 9:05 Bell rings, S's start coming into room.
- 9:08 T comes into room.
- 9:11 T takes hot lunch and milk count and attendance.
- 9:15 T dismisses S's and all but 23, 26, 19 (12 is absent) leave and go to reading and spelling in other rooms.
- 9:16 S's from 3 other rooms come in LL room for reading and spelling. T chats with observer, gets books arranged for reading, puts file folders and workbooks out on TD#1.
- 9:20 P. 92, 92 -- Pink; p. 116 -- Blue; p. 108 -- Purple. The T tells each of the 3 groups that these are the pages to write their spelling words on.
- 9:21 T starts giving dictation from spelling workbooks; he starts giving dictation to the Pink group 1st, then dictates 2 words to the Blue group, and then dictates 2 words to the Purple group. He dictates 2 words at a time to each group and then says a sentence to each group which contains both words just dictated. T has moved TD#1 up to the NW corner of the room just in front of the CB. TD#1 is a waist high table with wheels on it. On top of TD#1, the T has TE of TB and workbooks arranged. While giving dictation, the T either walks over to the group or faces the group to which he dictates the words. In all 20(?) words are dictated to each group. HR S's in reading/spelling class:
Jill - 23
Vickie MC - 26
Lee Ann M (absent) - 12
Tonia Y - 19
- 9:35 Another T came into the room and talked to LL for about 30 seconds. S's were writing words at this time.

- 9:39 Dictation stops; T asks S's to correct words, he waits for room to quiet down.
- 9:40 T goes to TD#1 and looks at some materials on the desk.
- 9:41 T -- "Complete last week's dittos. Leave your workbooks with me. We have to complete the skills unit. Finish vocabulary and then start on TB study." This assignment is for the Pink group. They begin working independently.
- 9:42 T dictates a sentence to S's in the Blue group.
- 9:43 T asks S to repeat the sentence.
- 9:44 T dictates another sentence. He dictates each sentence only once. S's are expected to be able to write down correctly the whole sentence dictated. He calls it a "fluency program."
- 9:46 T asks S to read the sentence. He then asks S to read what she has written.
- 9:47 T -- "... in Skills Reader, we're going thru the TB study to emphasize important parts. Turn to p. 243 in Skills Reader ... on p. 242." T reads directions out loud. T then allows time for S's to read the selection silently. This is the Blue group he is now working with. Pink (T sometimes calls them Orange group) group working independently.
- 9:50 T stands at TD#1 waiting for S's to read the selection from p. 242 of Skills Reader.
- 9:51 T walks around helping S's who are having trouble deciding on a title for the selection.
- 9:52 T goes to CB and writes:
- I.
 - A.
 - B.
 - II.
 - A.
 - B.
 - III.
 - A.
 - B.
- 9:53 T asks S what the title is. T writes on the CB what he says: Why the South Has Busy Industry. T asks the group what the main topics of the

selection are. S's answer and T writes them on the CB.

- I. People
- II. Water
- III. Natural Resources

9:54 T asks for 2 important details for each main topic. S's answer and T writes them on the CB. T helps S's narrow their responses to one or two words.

I.

- A. Manpower
- B. Demand of products

- II. A. Mountains streams S's felt there
- B. Powerful rivers should be 3 impor-
- C. Electric power tant details under
- II.

10:00

III.

- A. Mineral deposits
- B. Raw materials

10:02 T -- "For you S's who had finished this assignment, is this how your outline looked?"
T -- "How many have completed their reading papers?"
T begins passing out papers that S's had completed earlier. T had corrected them. T asks S's to correct their errors on the papers.

10:03 T completes handing out papers. Several S's bring up papers to him. He looks them over.

10:04 T gives directions to Blue group S's for reviewing of paragraph skills and ... skills and Vocab. review.

10:05 S's ask questions, T answers.

10:07 T walks around Blue group checking to see if S's are getting started correctly.

10:08 T stops and explains to the Blue group that on ditto, 43, question 20, only 1 S got this correct. T -- "What does percolating mean?"

10:09 T walks around the Pink group to see how S's in the Pink group are working.

10:10 T goes back to TD#1. S comes up and asks for help. T helps the S.

10:12 T goes to bookcase by windows and takes out some paperback books and gives them to a HR S to read.

The books are Readers Digest skill books. (This is the S who is on an individualized reading program. T calls her the HR group.)

- 10:13 T goes back to TD#1, S comes up and asks for help. T helps the S.
- 10:14 T walks around helping any S in the CR who needs help. All S's are working independently in TB and workbooks. 1 S is working in a Readers Digest book. 19 is looking at a globe. Some S's are working on a packet of dittos, #38-41, and 43 from ditto master for "Reading to Learn." Jill-23 wanders around the room from 10:23-10:30, Jill is in both Purple and Blue spelling groups.
- 10:29 T -- "Tomorrow we will work in Barnell-Loft and SRA. Those who have not finished work in the unit can have time to do so." T talks to the whole group.
- 10:30 T -- "Alright, that will be all for today." S's start leaving room, HRS, #2 - 18, 20-22, and 24, 25 start coming back into the room.
- 10:31 Transition.
- 10:34 Recess.
- 11:02 S's start coming into room.
- 11:07 T -- "Today we are going to review the addition of unlike fraction." T passes out dittos to the whole group, #2-11, 13-26; T tells S's to begin working on them because he needs a few minutes to prepare metric materials for the class. S's begin working on ditto. T walks around room getting materials ready. S's have trouble on the ditto, so T interrupts his preparation.
- 11:09 T goes to CB and writes:
- | | | |
|------------|--------------|--------------------|
| $1/3=4/12$ | 3,6,12,15... | $1/2$ |
| $1/4=3/12$ | 4,8,12,16... | $1/7 \quad _ /14$ |
- S's have trouble remembering how to find common denominator. (CD) T explains how to find CD, multiples of each denominator the common multiple (CM) will be the CD.
- 11:12 T walks around helping S's find common denominator.
- 11:17 T leaves room.

11:22 T returns, T -- "It has taken me longer than I had anticipated, so we'll work on that tomorrow."

11:23 T -- "Those of you who have finished can look ahead in your metric book. Those having trouble, look up here." T does common denominator problem on the CB. It is problem #2 on the ditto.
 $2/3 = _ / 24$ (written on board).

$$1/8 = _ / 24 \quad \text{whole number } \times 8 = 16, 24$$

11:26 T walks around to see if S's are getting correct answers, then back to the CB to complete problem.

$$\begin{array}{r} 2/3 = 16/24 \\ + 1/8 = 3/24 \\ \hline 19/24 \end{array} \quad \begin{array}{l} \text{T asks S's to tell him what} \\ \text{the answer is.} \end{array}$$

T -- "You should ask yourself, can this be reduced to lowest terms?" T asks 20 for the answer.

11:28 T -- "Look at problem 3." T puts this problem on the CB.

$$\begin{array}{r} 2/5 = 8/20 \\ + 1/4 = 5/20 \\ \hline 12/20 \end{array}$$

T -- "Say to yourself, you multiply the denominator by a series of whole numbers."

T -- "Can the problem be reduced? What do you say to yourself, 20?" S -- "13 is prime so you can't reduce it."

11:30 T -- "Problem 4." T puts it on the CB.

$$\begin{array}{r} 3/4 = _ / 20 \\ + 5/5 = _ / 20 \\ \hline \end{array}$$

T -- "25, what is the CD?" S -- "...", S -- "20"

11:32 T -- "We have time for 1 more." T puts this problem on the CB.

$$\begin{array}{r} 5/6 = 15/18 \\ + 2/9 = 4/18 \\ \hline 19/18 \end{array} \quad 1 \frac{1}{18}$$

T -- "This fraction is > 1 ." "How much greater?" He answers his own question.

11:33 Transition.

- 11:35 S's begin to leave for lunch.
- 12:10 Bell rings and S's start coming into room. They take out books and read independently. USSR
- 12:13 T comes into room, chats briefly with observer and then takes roll. Several S's chat briefly with him.
- 12:17 T leaves room. 23 came over and sat down next to me. She does not want to read during USSR.
- 12:20 T returns, picks up 16mm film and then leaves room again. He sets up film on projector in the hall.
- 12:21 T returns, picks up TE of Language book and puts it on TD#1; hangs up his coat in the closet; puts extra math dittos away; answers some S's question; gets AV order blank out of his briefcase and then sits at a table and begins filling it out. S's continue to read independently.
- 12:26 7 leaves to go with reading specialist.
- 12:30 T -- "Put away your work and go to science." Transition. T stops working on order form. He pulls down film screen and closes curtains. S's 2 - 26 leave for science. T brings in the film projector and sets it up in the SE corner of room.
- 12:32 S's from another room come in.
- 12:34 T talking to the whole group -- "The film is an introduction to U.S. geography and land forms." (The film is a Cornet Film of about 1953 vintage.)
- 12:35 T starts 16mm film, then adjusts curtains, gets record book from TD#1 and then sits by the projector working on record sheets and watching the film. Topics covered in the film:
- Mineral deposits -- oil, ore
- Great Seal of the US and motto -- E Pluribus Unum
- Belts -- Corn; wheat; cotton, tobacco, oats, soybeans, hay; corn belt one of the most important regions. Corn for humans and for livestock. One of the richest regions in the U.S., importance of the products grown for the economy and jobs.
- Grazing and irrigated crops region -- mountains and high plateau; sheep and cattle in this region, need of water in this region. Dams built to store

water allows for irrigation. People live in cities because of variety of jobs. 2/3 of 180,000,000 people live in cities. How are these people linked together? The 12 largest cities are near water. Great Lakes -- St. Lawrence Waterway. Ohio, Mississippi and Missouri Rivers used for transportation. Ore sent over the GL and St. Law. Waterway. Waterways tie people, natural resources and products together. Railroads do the same.

Oil pipelines, air transportation tie people and regions together; transportation links people, resources and commerce together. It ties U.S. people together; highways are especially useful for people to be tied together.

- 12:47 Film ends. T talks to the whole group -- "Did you find any new information in this film?" S -- "About St. Law. Seaway" ... "mining pits."
- 12:48 T led discussion with whole group, T passes
- 12:50 around a chunk of natural copper ore. Discussion was on ore deposits and ore ranges. T uses map of U.S. to point out where copper and iron ore ranges are located. In northern MI, WIS, and MINN. Points out how the ranges extend into Canada. Tells about his brother-in-law who works in a copper mine. 7 returns.
- 12:53 T passes out taconite ore pellets for S's to handle and look at. T explains how taconite pellets are made. Use low grade ore; ore tumble in large magnetic drums.
- 12:54 T passes out chunks of natural iron ore just as it came from the ground. Points out natural orange color.
- 12:55 T shows different chunks of copper ore from different ranges in the U.P. Some are solid copper, others have impurities mixed in.
- 12:56 T starts to collect the ore samples from the S's.
- 12:58 T explains about the St. Lawrence Seaway. He points it out on the U.S. map.
- 1:01 S says he was surprised to see that the Hawaiian Islands were so far north and he wanted to know why it wasn't cold there. T explains the difference in climate.
- 1:05 S's leave -- transition.

- 1:08 T explains that the film is an introduction to the geography of the U.S. -- land forms, climate, transportation, crops.
- 1:10 S's 20 and 3 run the film -- same film as was shown to the 12:30-1:00 group. Whole group watches; additional topic on the film: variety of climates -- helped to make the U.S. the leading agricultural nation of the world. T watches film.
- 1:17 T stops film 7 minutes into it and asks S's to recall names of the great rivers of the U.S. -- Ohio, Missouri, Mississippi. He points them out on the map of the U.S. hanging on the front CB.
- T -- "How many dams on the Colorado River? ... about 23." 6 asks a question on what would happen if one of the dams broke, T explained that it would depend on how much water was behind the dam. In other words, the time of year would help determine how much water.
- 1:21 T talked briefly about the earthen dams that had recently broken. Several more S's question on dams and floods. (T had to justify morally the use of the word dam.)
- 1:22 22 comes into the room from science; the dams and floods discussion led a S to relate experience they had recently with severe weather conditions.
- 1:25 22 leaves the room.
- 1:27 T points out major cities on the U.S. map.
- 1:28 T starts film again and sits and watches; additional topic on film: kinds of products carried by rail -- cattle, sand, gravel, lumber, oil.
- 1:31 22 returns.
- 1:33 T writes 180,000,000 on the CB.
- 1:34 Film ends; 6 leaves room. T asks S's to relate things they learned from the film -- S's responses -- St. Lawrence Seaway, dams.
- 1:35 3, 22 leave room.
- 1:37 6 returns.
- 1:38 T talks about the Welland Canal. T told about seeing a Danish ship at dock in Deluth. 3, 22 return.

- 1:39 T talked about the air quality in Gary, Ind. T wrote EPA on the CB and then explained what it meant. Gary receives raw materials by boat.
- 1:40 20 goes to the U.S. map and points out where he went on a train trip and saw a lot of air pollution near Chicago. He wondered why nothing had been done about it.
- 1:42 10 leaves the room.
- 1:43 T asks S's if they could tell from the cars shown on the film when the film was made, also the tractors and harvesting equipment. 11 leaves the room.
- 1:44 10 returns, 11 returns.
- 1:45 Principal comes into the room and talks briefly to T.
- 1:46 S's decided the film was made in the '50's. S's related information they had about tractors and equipment used to build roads. S's talked about family and neighbors who had tractors and equipment they had seen on construction sites.
- 1:48 T tells S's to take out English books and turn to chapter 6, p. 169.
- T -- "Here is listed what we are about to study."
- 1:49 T -- "question 1: What is jargon?" T -- "You people who are still having trouble finding the page, turn to the table of contents and see what page chapter 6 starts on."
- T -- "question 2."
- 1:51 T opens curtains.
- 1:52 Question 3) T continues to briefly introduce the topics listed on p. 169, he briefly explains what each topic is. He sometimes ties a topic in with what had been previously studied. Topics which tied in with previous study were 3); 4); 7).
- 2:00 For sentence 12) T asked S's to relate a riddle. S's 6, 16 and 24 gave riddles.
- 2:01 20 said one could advertise kittens free in the State Journal. T asked why. 20 said maybe the Humane Society arranged it.

- 2:04 Secretary brings in notes and tells T they have to go home after school.
- 2:05 T asks 23 to write S's names on notes to send home; she sits at a table on the side of the room putting names on notes.
- 2:06 T -- "O.K., let's take a look at jargon (topic 1)." T asks 24 to read aloud to the class from p. 170.
- 2:07 T tries to get S's to come up with the word caption. They don't so he tells them about caption.
- 2:08 24 reads the next question from p. 170. T discusses jargon words, swabbing and bulkhead.
- 2:10 24 reads orally from p. 170, T explains meaning of "jargon" in terms of swabbing and bulkhead -- words used in the Navy.
- 2:12 24 reads orally from p. 170. T asks 11 to respond, 22 leaves the room.
- 2:13 24 reads orally from p. 170. 22 returns.
- 2:14 T asks 6 and 22 if police have jargon, (their dads are policemen). They said they think so.
- 2:15 T tells about visiting the FBI headquarters in Washington D.C.
- 2:17 2 reads orally from p. 170.
- 2:18 T asks S's to discuss meaning of jargon in sentences 1) and 2) on p. 170. 23 finishes writing names on notes.
- 2:20 24 reads orally last sentence on p. 170.
- 2:21 10 reads paragraph at top of p. 171; T and S's discuss meanings of jargon in sentences 1-4) on p. 171. 3 reads question 2.
- 2:24 23 begins writing student names on another set of notes to go home, S-23 decided to do this on her own.
- 2:26 T tells jargon from golf that is now little used -- brassey, mashey, spoon. S's had never heard these words!
- 2:27 8 reads 3) out loud on p. 171.

- 2:29 26 reads orally from "For Practice" on p. 171.
- 2:30 14 reads orally 1) under "Written" on p. 171. T asks S's if they watch "Quincy" on t.v. and briefly discussed "morgue" in this context.
- 2:32 6 looks up morgue in dictionary.
- 2:33 ... reads 2) orally, p. 171; T was in a radio station once when it was cut off the air. It took about 3 minutes to build power back up in order to put the station back on the air.
- 2:34 23 stops writing names on second set of notes.
- 2:35 21 reads 3) orally, p. 171; T -- "Havlichek started playing BB in 1962. Yesterday he got a standing ovation.
- 2:38 6 stops looking for morgue in dict. T reads the definition out loud to class.
- 2:39 25 reads 4) orally, p. 171.
- 2:40 T asks 3 to review jargon.
- 2:41 T -- "Preview for tomorrow" p. 172. T asks 11 to comment on it. P. 173 T asks 15 to comment on it. P. 174 T comments and so does S-19.
- 2:45 S's leave for recess; T goes outside to supervise recess.
- 3:10 S's begin coming in from recess cause it's raining; S's get books, coats and misc. materials ready to go home and also "play" in the room.
- 3:15 T exchanges small talk with S and then gets safety patrols organized to have "rainy day bus line-up."
- 3:20 S's dismissed to go home.

APPENDIX B

APPENDIX B

Coding Procedures

General Procedures

1. Each student in each class is assigned a number (01, 10, 20) which remains constant for all coding procedures.
2. Class refers to a number assigned to each teacher in the study.
3. Day refers to the date of the data source which is coded.
4. Source indicates whether teacher logs, observations, or teachers plans were used as the data source.
5. Beginning and ending times refer to the time activities started and stopped.
6. Subject areas include types of activities found in school days with provisions for major and minor areas.
7. Group refers to whole group, subgroup or individual.
8. Group size refers to the number of students in the group considered; check attendance to determine group size.
9. Supervisory code refers to teacher supervised, other supervised or nonsupervised.
10. Location refers to in own room, out of own room, or out of school.
11. Process variable refers to the amount of actual reading or writing done by students during a time interval. Writing refers to text or sentence compositions, not to penmanship.

Student Procedures

12. Use the same subject numbers throughout all of the coding for a given classroom. That is to say, subject 25 must refer to the same person in all of the coding.
13. If a child is absent record on the code sheets his number, class, day, and source. For the beginning time, give the beginning time for all other students for that day, and for the ending time, use the ending time for that day. Be sure to check attendance and note those children that are absent on the code sheets.
14. If some pupils are not identified ignore their actions in the coding or if they are identified but only as involved in momentary actions, ignore them in coding (anything 30 seconds or less or "brief" is defined as momentary).
15. If a beginning and an ending time cannot be found for children leaving the room, ignore their having left, i.e. treat them as if they never left the room.

Time Interval Procedures

16. Times for intervals must be continuous, e.g. 9:12 - 9:20; next interval 9:20 - 9:40; next 9:40 - .

Subject Area Procedures - General

17. Always consider the large unit when classifying subject areas. If a larger segment of time which is homogenous with respect to content has embedded in it only a short comment by the teacher which would change the content specification, ignore this comment and code for the larger unit.
18. When the teacher gives directions or elaborates on an assignment, this should be coded in whatever subject area it occurs; it is part of the time interval for the subject area coded.
19. Announcement of due dates should never be coded separately.
 - a. If due dates are announced during a regular lesson, then treat the announcement as part of the subject area in which it occurs.
 - b. If due dates are announced during a transition,

consider the announcement as part of the transition.

20. For the codes 0100, 0200 and 1500 no minor is usually coded.
21. When children leave for the library, code the content of what they will be doing in the library if you know it. If children leave during a period in which they were instructed to use the library as a resource, then code their subject area as the same as what the rest of the class is studying during that time interval: only code them being out of their room by location code. If children leave during some other time when the content is not clear, or during the reading or language arts period, or during their free time, assume they have gone to pick out a library book for their free reading time: code these students as 0212 and 12 on the process variable.
22. There is no separate code for tests. All testing should simply be coded as to the subject matter which it covers. For the supervisory code, code it as 1 - teacher supervised. For the group designation code, code it as individual. For the process variable code - code it as 30.
23. Code movies or tests or field trips or educational assemblies in terms of the content involved for subject area.

Subject Areas Procedures - Language Arts

24. Code all sharing activities as 0110: Language Arts - oral communication. If children spend time with speech and/or hearing therapy code them as 0110.
25. Writing instruction under language arts includes instruction in the process and art of writing as well as structured practice in writing; it does not refer to penmanship.
26. Sentence composition refers to composing sentences only - not to text composition. Sentence completion is sentence composition if it involves more than one word.
27. If as a part of the language arts lesson children are taught to read maps, tables, or graphs, or to develop map legends, tables or graphs, this should be classified as 0180 - information gathering

skills.

28. The category "literary forms" under language arts is for content dealing with various literary forms such as poetry, autobiographies, biographies, fairy tales, folktales, and tall tales. If the reading lesson aims at reading literary forms then "literary forms" should be used as the minor designation.

Subject Areas Procedures - Language Arts and Reading

29. For reading and language arts use the teacher's specification (from the schedule or the blackboard or convention) as to whether the major code is reading or language arts.
30. For all reading and language arts lessons where the major specification is reading or language arts, code the content of the reading, writing, spelling, etc. lesson as the minor content specification. If the content does not fit one of the codes, such as science, social studies, etc., then and only then leave the minor code blank. Do not stretch the point in coding the minor area. In a fairly straightforward way, it must be science, social studies, etc. before it is coded as such.
 - a. Reading lessons can have a minor in language arts, and vice versa.

Subject Area Procedure - Reading

31. The reading categories are defined as follows:
 - a. No explicit analysis - no overt attempt is made at analyzing what is read.
 - b. Word analysis - includes phonetic analysis, structural analysis and sight words.
 - c. Word meaning - vocabulary development.
 - d. Text analysis - comprehension, sequencing, main events, main idea, setting, etc.
 - e. Individual reading - child is reading by himself either silently or to the teacher.
 - f. Group reading - the activity where a subgroup meets with the teacher and some or all of the children alternate in reading the text and sometimes answer questions about what they have read. Also - where questions are asked and the children then read silently to find the answers. To be coded here, children must be reading. If the children are reading paragraphs

- from their workbooks in class with the teacher and then discussing them, code this as group reading.
- g. Lecture or discussion - where the teacher lectures on or the teacher and students have a discussion about reading itself. Also - for situations where there is a discussion about the content of what has been read but there is no reading - (either silently or out loud) - during that lesson. Also - when the teacher lectures (talks) about reading, word analysis, or literary forms without actual reading by the students.
 - h. Individual reading and doing exercises - where the child reads by himself/herself and does exercises based on the reading.
 - i. Doing exercises (dittos, tapes) - where children are doing only the exercises. If the teacher is discussing their answers with them, this is coded as discussion. If an individual child reads with T and they discuss the text this is coded as individual reading with T's supervision: don't code as discussion.
32. If more than one of the reading levels (on the third digit, e.g. word analysis, etc.) occurs during a lesson, code as follows:
- a. If the different areas are covered separately and are sequenced one after another and are of at least 2 minutes in duration, code the different parts separately. Create a new time interval for each part of the lesson.
 - b. If the different levels are distinct and sequenced but short in duration (all but one less than 2 minutes), code the whole lesson as one time interval and code it hierarchically, giving the level with the highest code the greatest priority (e.g., if both word analysis and text analysis occur and word analysis is less than 2 minutes in length code the whole interval as text analysis).
 - c. If the different levels are intermixed in the lesson, code the whole lesson as one interval and use the level with the highest code.
33. If a child is reading with an aide, classify the subject matter as 0200 and code the process variable as 12.
34. If a child is doing a crossword puzzle and it is not clear from the context that the purpose of it is for word analysis, then code 3 in the third digit for reading, or word meaning.

35. In reading on the 4th digit (individual reading, group reading, etc.) make a new interval for activity change and code it separately. Do not code the whole lesson or use the notion of an hierarchy.
36. When dealing with reading groups, code the children involved in that reading group 0900 from the moment the teacher calls them up for the reading group until the point at which the actual instruction begins. When the children finish with the reading group and are dismissed, code them 0900 from that point until the point at which it is recognized that they have actually begun work on some other matter. If this is not indicated, then do not code them as 0900 but simply code them as having returned to seatwork or whatever else it is that they are doing. This latter case will most likely be prevalent.

Subject Area Procedures - Social Studies and Science

37. The distinction between Social Studies and Science revolves around the focus of content. If the focus is technical, then it is science. If on the other hand, the focus is on the effect that some scientific or technological field has on society or individuals, then it is coded as social studies.
38. If during a science or social studies lesson the teacher instructs the students in reading or some area of language arts, be sure to code reading or language arts as the appropriate minor. To be coded as a minor instead of as a process (see convention 60) there must be formal instruction or formal feedback in the area.
39. Social Studies includes history, geography, sociology, anthropology, government, political science and economics (all coded as 0800). Lessons dealing with social behavior and affective goals and values should be coded as Social Studies (0810).
40. In terms of the code 0810 (subject area), only code lessons where there is formal instruction in the area of values or social attitudes. Do not code momentary interactions about values, behavior in the classroom or issues of discipline under the code 0810.
41. "Child of the week" is coded 0810.

Subject Area Procedures - Breaks, Beginnings, and Endings

42. Codes 09-13 for Subject Areas indicate various breaks.
 - a. 09 - for between instructional activities including the passing out or collecting of materials. If a child spends time with a social worker code him as 0900.
 - b. 10 - only for recess or lunch.
 - c. 11 - all activities at beginning or ending of day (or half of the day) including lunch money, clean-up.
 - d. 12 - if children disappear for short periods of time from their room and it is not clear where they went code them as 1200.
 - e. 13 - any other break such as fire or tornado drills; other people enter the room, etc.
 - f. If children come in late at the beginning of the day, code them as 0900 until they arrive; if they come in late after lunch code them as 1000 until they arrive.
43. Whatever happens at the beginning of the day or at the beginning of the second half of the day (before the teacher formally begins the activities) is coded 0900 or transition. When the teacher begins, this could be coded as 1100 if it is a beginning or ending exercise or as the regular subject matter if there are no beginning exercises.
44. For transitions or breaks do not code process variable, group, supervisory code location, etc. Just code times and break code.
45. Children leaving and returning during transitions or breaks or opening exercises need not be separately recorded, as long as they leave and return during the break.
46. For transitions to and from reading subgroups, code them for the children involved when the information is available. For the beginning of the group lesson code the transition from the time the teacher announces the group to the class to the time at which the lesson begins with these children. If there is confusion as to the beginning time vs. the transition, code the lesson as having begun immediately. The end of the subgroup comes when the teacher announces they are finished. If there is no further reference to these children returning to their seats or beginning other activities, code without transition.

47. Make a judgment about when the transition is over using the criterion of when most children have begun to work.
48. Code all passing out of materials as transitions.

Subject Area Procedures - Seatwork

49. If the child is doing seatwork during the reading lesson and is reading in his reader, and it is not clear to the rater whether the reading instruction is aimed at word meaning, text comprehension, or whatever, classify subject as 0202. The third digit 0 means that it could not be ascertained what the nature of the reading task is, but it is known that the child is working in reading and he is also doing reading by himself. Likewise, if the child is working on some ditto or a workbook and if it cannot be ascertained from the observation what the exact nature of the exercise is, classify as 0203. This indicates he is reading and working on exercises, without knowledge of the exact nature of the material. If you know whether it is word meaning or text analysis or whatever, then, of course, code this in the third digit. If the child is intermixing the two, that is, reading and also doing exercises and it is not possible from the observations to know at what point the child stopped reading and began doing the exercises based on that reading, then use the code 6 in the fourth digit for reading. This indicates both reading and exercises are being done during that period.
50. If during an individual work period the teacher makes an announcement about the fact that the children ought to move on to task B when finished with task A the fact that both A and B are now possible tasks must be accounted for. This will usually necessitate the use of mixed seatwork code 15 with the third digit indicating, if it is possible, which two subject matters are being included in the mixed seatwork. However, if the teacher does not change subject matters by her announcement; that is, both assignments are in reading, or both assignments are in language arts, then there is no need to move to the 1500 code.

Group Designation Procedures

51. Group designation refers to the nature of the instructional setting.

52.
 - a. For group designation, if more than one child is involved, but less than the whole class, code as subgroup.
 - b. Movies and assemblies are whole group activities unless otherwise specified.
53. Code the group size variable for all intervals but do not change it to reflect momentary changes in group size such as toilet, library, etc., breaks for individual children in the group.
 - a. For group size involving standard groups just take the given number in the group minus those children that are absent for that day.
 - b. For all non-standard groups count the number involved.

Supervisory Code Procedures

54. For the supervisory code, it should be coded teacher or other supervised only if the teacher or aide is actively involved in educational supervision or monitoring of student activities.
 - a. If a teacher is walking around the room and supervising seatwork by interacting with the children and all the interactions are momentary, code all children during this period as having been supervised.
 - b. If the teacher is at his/her desk or is walking around and has a 30 second or longer interaction with a child, code the child as having been supervised during this interval and all other children during this interval as not having been supervised.
 - c. If the teacher is at his/her desk or a table working on something by his/herself or watching the children, and children come up to the teacher for momentary interactions, code all children during this interval as non-supervised.
 - d. All whole group or subgroup teacher instruction is coded teacher supervised for the children involved.
 - e. Code the showing of movies, instructional use of tapes, records, etc., typically as supervised.
55. Only use the category "other supervised" when it is some individual other than the teacher, such as an aide or another student who is used as an aide in the classroom. If the children leave the room and receive their instruction from the music teacher, the P. E. teacher, or the art teacher, code them

as having been teacher supervised. Also code children during the time they are in the library as teacher supervised (unless there is no person formally assigned as a librarian).

56. If a child is near the teacher, working by himself/herself and the teacher is also working by himself/herself, the supervisory code is 3 - nonsupervised: close physical proximity to the teacher does not count as supervision.
57. If during an observation a child is recorded as having come up to the teacher for instruction and the next instance recorded is of new child being called up to the front, at that point (unless otherwise specified in the observation) code the other child as having returned to his seat. Most observations should indicate both the time they came up and the time they returned to their seat, but if not, use the above convention.
58. Do not forget that when the supervisory code changes, i.e. the teacher starts or stops to actively monitor, a separate time interval has to be coded.
59. Ignore any individual discipline problems in the classroom, no matter the length of time involved, unless they interrupt the teacher while he/she is with some other children who are receiving instruction. The point is that the interaction must take teacher time or supervision away from other children.

Process Variable Procedures

60. For the process variable, code whether during the time interval in question the student himself/herself did any writing or reading. The student must have actually done the reading or writing. If both occur, code it 4.

The process variable records the act of reading or creative writing, not formal instruction in reading or writing, which is recorded as a major or minor. The second digit records roughly what proportion of the interval was spent in reading or writing. Reading must involve more than reading directions or sentences on a ditto - it must involve the reading of text. Writing is also classified only when the child writes text - not merely filling in words on dittos or copying material from the board. In those instances where the child makes up

a story but does not write it himself, this is not classified as writing. To be classified as writing more than a sentence must be involved. If instruction in writing is provided but the children do not actually write themselves, code the major as 0170, and the process variable as 30.

61. For the process variable; if the children leave the room to go to a reading class with another teacher, code them 12.
62. For the process variable; code children working in their workbooks as 30.
63. If there is no information about the process variable (reading and/or writing) which can be broken down to the individual level for time intervals, code the process variable 00.
64. For USSR; code 12 for the process variable (after transition, if applicable); USSR represents a structured opportunity to read.

APPENDIX C

APPENDIX C

ANOMALIES IN MATCHING OBSERVATION AND LOG PURSUIT RECORDS

12/17/81

Student 23 in class 3 on 4/27 was dropped from pursuit analysis due to the fact he/she was coded absent by the observer and not by the teacher.

12/29/81

Student 12 in class 4 on 5/12 was dropped from pursuit analysis due to the fact he/she was coded absent by the teacher and not by observer.

1/18/82

On student 2 in class 6 on 5/10 the observer coded 10, 0, 0 from 11:35 to 1:15 while the teacher coded a variety of activities listed below:

minutes	major	group	sup
5	1	1	1
25	15	3	1
5	0	0	0
35	10	0	0
3	9	0	0
22	15	3	1

This problem was handled by matching the observation pursuit with the 35 minute 10,0,0 log pursuit.

1/21/82

On student 19 in class 6 on 5/10 the observer coded 83 minutes of 5, 3, 2 from 1:52 to 3:15 while the teacher coded a variety of activities listed below:

minutes	major	group	sup
8	15	3	1
3	9	0	0
12	2	1	0
20	7	1	1
5	3	3	0
5	9	0	0
25	4	1	1

This problem was handled by matching the observation pursuit with the 25 minutes of 4,1,1 since it was the largest block of time.

1/21/82

On student 2 in class 6 on 5/25 the observer coded as minutes of 13, 0, 0 from 1:49 to 3:15 while the teacher coded a variety of activities listed below:

minutes	major	group	sup
11	2	3	1
15	10	0	0
5	9	0	0
10	2	1	1
40	7	1	1
5	9	0	0

This problem was handled by matching the observation pursuit with the 40 minutes of 7,1,1 since it was the largest block of time.

1/27/82

On student 2 in class 6 on 6/06 the observer coded 70 minutes of 5, 3, 3 from 12:50 to 2:00 while the teacher coded 35 minutes of 5, 3, 1 and 30 minutes of 2, 3, 1 and 5 minutes of 9, 0, 0 during that time period. The observer pursuit record was matched with the 35 minutes of 5, 3, 1.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Allington, R. L. Poor readers don't get to read much. Occasional paper 31, The Institute for Research on Teaching, Michigan State University, East Lansing, MI, 1980.
- Arlin, M., Roth, G. Pupils' use of time while reading comics and books. American Educational Research Journal, 1978, 15 (2), 201-216.
- Bennett, N. Teaching style and pupil progress. London: Open Books, 1976.
- Bloom, B. S. Human Characteristics and School Learning. New York: McGraw-Hill, 1976.
- Bock, R. D. Multivariate Statistical Methods in Behavioral Research. New York: McGraw-Hill, 1975, 507-559.
- Borg, W. R. Time and school learning. In Denham, C. and Lieberman A. (Eds.), Time To Learn. National Institute of Education. May 1980.
- Brophy, J. E. Teacher behavior and its effects. Journal of Educational Psychology. 1979, 71 (6), 733-760.
- Carol, John B. A model of school learning. Teacher's College Record. 1963, 64, 723-733.
- Colin, H. M. Accuracy of teacher reports of their classroom behavior. Review of Educational Research. 1979, 49 (1), 1-12.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. The Dependability Behavioral Measurements: The Theory of Generalizability for Scores and Profiles. New York: John Wiley and Sons, 1972.
- Ebmeier, H. H. & Ziomek, R. L. Engagement Rates as a Function of Subject Area, Grade Level and Time of Day. Paper presentation, American Educational Research Association Annual Meeting, New York, April 1982.
- Fisher, Charles W. A study of instructional time in grade 2 reading. San Francisco, Ca.: Technical Report

II-4, Beginning Teacher Evaluation Study, Far West Laboratory for Educational Research and Development, November 1976. ERIC Document Reproduction Service No. ED 145 414.

Fischer, C. W., Filby, N. N., Marliave, R., Cahen, L. S., Dishaw, M. M., More, J. E., and Berliner, D. C. Teaching behaviors, academic learning time and student achievement. Final report of Phase III-B. Beginning Teacher Evaluation Study (Technical Report V-1). Washington, D.C.: National Institute of Education, June 1978.

Fredrick, W. C. & Walberg, H. J., Rasher, S. P. Time, teacher comments, and achievement in urban high schools. The Journal of Educational Research. 1979, 73 (2), 63-65.

Fredrick, W. C., & Walberg, H. J. Learning as a function of time. The Journal of Educational Research. 1980, 73, 183-194.

Harnishchfeger, A., Wiley, D. E. The teaching-learning process in elementary schools. A synoptic view. Curriculum Inquiry. 1976, 6 (1), 5-43.

Karweit, N. A reanalysis of the effect of quantity of schooling on achievement. Sociology of Education. 1976, 49, July, 236-246.

Karweit, N. Quantity of schooling: a major educational factor? The Educational Researcher. Feb. 1976, 15-17.

Karweit, N., & Slavin, R. E. Measurement and modeling choices in studies of time and learning. American Educational Research Journal. 1981, 18 (2), 157-171.

Millman, J. & Glass, G. Rules of thumb for the ANOVA table. Journal of Educational Measurement, 1967, 4, 41-51.

Schmidt, W. H. Measurement error: Should it include a specification for bias. Paper presentation. American Educational Research Association Annual Meeting, Los Angeles, April 1981.

Schmidt, W. H. The high school curriculum: It does make a difference. Curriculum Inquiry, in press.

Shavelson, R., Dempsey-Atwood, N. Generalizability of measures of teaching behavior. Review of Educational Research, 1976, 46 (4), 553-611.

Stallings, J. Allocated academic learning time revisited, or beyond time on task. Educational Researcher. 1980, 7 (11), 11-16.

Symonds, P. The correlation of English with other subjects from the point of view of psychology. The Elementary English Review, 1930, 7 (1).

Wiley, D. E. Another hour, another day: Quantity of schooling, a potent path for policy. In Sewell, W. H., Hauser, R. M. & Featherman, D. L. (Eds.), Schooling and Achievement in American Society. New York: Academic Press, 1976.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03175 1740