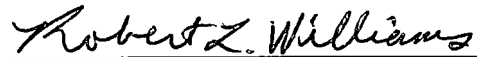


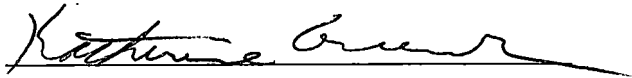
To the Graduate Council:

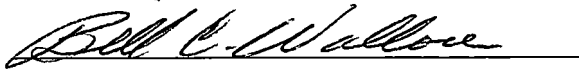
I am submitting herewith a dissertation written by Eun Jung Oh entitled "The Psychometric Properties of the Student Work Habits Rating Form." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.



Robert L. Williams, Major Professor

We have read this dissertation
and recommend its acceptance:







Accepted for the Council:



Associate Vice Chancellor and
Dean of The Graduate School

**THE PSYCHOMETRIC PROPERTIES
OF THE STUDENT WORK HABITS RATING FORM**

**A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville**

**Eun Jung Oh
May 2000**

Copyright © Eun Jung Oh, 2000

All rights reserved

ACKNOWLEDGMENTS

There are many people who helped me complete this dissertation. I would first like to thank my Dissertation Committee, Robert Williams, Bill Wallace, Ralph Brockett, and Katherine Greenberg, for their support and advice during the completion of this study. I especially thank my Chairperson, Robert Williams. I appreciate his dedication to reading and editing many revisions of my dissertation and all his time and effort.

I am grateful to Soo-Hee Park for his help with the statistical work of this study. I would like to recognize my family and friends for many years of encouragement. My father, Dae-Young Oh, and my mother, Hwa-Ja Jun, have especially been role models in my life. Their support and unconditional love made it possible for me to reach this moment. My friends who kept telling me that I could do it are Amanda Monville, Stephen Worth, James Canipe, Susan Stockdale, Barry Cox, and Larissa Chuprina. I am very lucky to have you all, and I will never forget you.

Finally, I would like to thank my fiancé, Jun-Hyung Park, for his love and understanding. I love you.

ABSTRACT

The purpose of this study was to evaluate the psychometric adequacy of a teacher rating scale of middle school students' work habits. This rating scale consisted of 20 work habits on which teachers rated selected students. A five-point Likert-type scale from *a lot like this student* to *very unlike this student* was used in rating the students on each work habit. Twenty-two teachers participated in the rating of 238 middle school students in 6th through 8th grades in 3 middle schools in the United States. Teachers rated the academic achievement of all students and the work habits of a subsample of students from each class. Teachers were asked to identify the 2 students with the best work habits and the 2 with the worst work habits in each class and then rate those 4 students on their specific work habits. The data analysis focused on the item contributions, internal consistency, construct validity, and criterion validity of the rating scale.

The main strength of the Student Work Habits Rating Form was its internal consistency, which was in the high .90s. This value greatly exceeds the .70 minimum traditionally required for research use. Thus, this rating form has considerable promise as a reliable means of assessing the work habits of middle school students.

Teacher ratings of student achievement were used in assessing the criterion-related validity of the Student Work Habits Rating Form. The work-habits ratings for the 3 achievement groups judged as poor, average, or outstanding were significantly

different. In particular, a logistic regression analysis showed that the work-habits ratings could classify correctly students in the poor and outstanding achievement groups virtually 100% of the time. Based on the result of the criterion-related validity of the Student Work Habits Rating Form, the rating scale appears to have substantial potential as a predictor of student performance.

The construct validity of the Student Work Habits Rating Form was assessed through a series of factor analyses, culminating in Confirmatory Factor Analysis. This series of analyses points to a general factor of work habits rather than a set of primary and secondary factors. This rating form provides an overall measure of work habits that might be a useful predictor of students' long-term success. In its current form, it should not be used as a diagnostic profile of areas where work habits are strong or weak.

Although this rating form appears to have considerable potential as a way to assess student work habits, all items on the rating form are not useful to an equivalent degree. Several conclusions can be reached from the various item analyses: (a) some items are more difficult than others with respect to receiving a favorable rating, (b) progression from low to high on the Likert-type scale is more difficult at the threshold from neutral to positive ratings than at the low end of the scale, and (c) most items discriminate best for students obtaining ratings toward the middle part of the work-habits distribution.

As a comprehensive rating scale, the Student Work Habits Rating Form has strong internal consistency and moderately good criterion-related validity. It provides a solid overall assessment of student work habits. Some items appear to be

psychometrically stronger than others, but all items (including the psychometrically weaker items) appear to have practical utility.

In spite of the psychometric and practical potential of the Student Work Habits Rating Form, several features of this study limit the generalizability of the results and the clinical utility of the rating scale: (a) the specific habits identified in the rating form did not reflect some of the important habits mentioned in the literature; (b) the student sample rated was too limited and unrepresentative of different ethnic groups; (c) too few teachers were used as raters, and their ratings were not cross-validated by other teacher raters; (d) no test-retest reliability assessment for the rating scale was done; and (e) more diverse and precise criterion measures need to be used in validating the scale.

TABLE OF CONTENTS

<u>CHAPTER</u>	<u>PAGE</u>
I. INTRODUCTION	1
Rationale	1
Statement of the Problem.....	6
Research Questions.....	7
II. REVIEW OF LITERATURE	8
Definition of Work Habits	8
Predictive Potential of Work Habits	10
Relationship of Grade Level to Work Habits	14
Strategies for Promoting Work Habits.....	15
Assessment of Work Habits.....	17
Development of Rating Scales.....	20
Conclusions of the Literature Review.....	23
III. METHOD	25
Subjects	25
Description of Instruments.....	26
Achievement Rating Form.....	28
Student Work Habits Rating Form	28
Procedures.....	29
Item Analysis	29
Reliability.....	29
Validity	30
IV. RESULTS	31
Item Analysis	31
Reliability.....	38
Validity Analysis	39
V. DISCUSSION	46
Findings for Research Questions	46
Psychometric Status of Student Work Habits Rating Form.....	47
Areas for Further Development	51
Conclusions.....	56
REFERENCES	57

APPENDICES	68
Appendix A: Achievement Rating Form	69
Appendix B: Student Work Habits Rating Form.....	71
Appendix C: Inter-Item Correlation Matrix.....	72
Appendix D: Parameter Estimates and Item- and Test-Information Functions.....	73
VITA	74

LIST OF TABLES

<u>TABLE</u>	<u>PAGE</u>
1. Demographic Information for the Work Habits Subsample From the Three Schools.....	27
2. Frequency Distribution, Mean, <i>SD</i> , <i>t</i> -value, and Effect Size for Each Item.....	32
3. Classical Item Statistics and Item Estimates From Samejima's Graded Response Model of IRT	35
4. Results for Criterion-Related Validity From the Known-Groups Approach.....	40
5. Results of Exploratory Factor Analysis: One- and Two-Factor Models	41
6. Confirmatory Factor Analysis Results: One- and Two-Factor Models	44

CHAPTER I

INTRODUCTION

Rationale

The purpose of schooling extends well beyond teaching the three Rs. Educators are now giving high priority to additional areas, such as social skills, emotional intelligence, and values. The development of student work habits should be added to this list of school priorities. Although relatively little research has addressed this area, good work habits could be among the most important contributing factors to students' success in school and as adults.

What is currently known about the work habits of American students? While some students are repeatedly disruptive and others consistently hard working, a majority of students have been characterized as passively compliant (Spaulding, 1992). They do the bare minimum, showing little initiative and enthusiasm about learning. This orientation to work undoubtedly limits what they will accomplish in school and may also severely impair their development as adult workers. Good work habits could be particularly important for urban students from economically disadvantaged backgrounds (Haberman, 1997). Without the educational and economic resources typically available to middle- and upper-class students, urban students must develop superior work habits to take advantage of educational opportunities and to increase their work opportunities as adults. Unfortunately, as Haberman points out, urban schools often model and reward antiwork behaviors.

Among the array of work habits likely to contribute to success in school and in the adult work force, one of the most fundamental is timely attendance at school. No student can take full advantage of school opportunities without being present on a regular basis. Absenteeism is especially damaging in academic programs requiring sequential learning (Ediger, 1987). Missing school establishes a vicious cycle that only worsens with time. As students miss school, they get behind in their work; and the quality of their work suffers. Poor performance is likely to make school more frustrating and less positively reinforcing, thus contributing to more absenteeism (Georgiady & Romano, 1994). This cycle is especially pronounced for special students (such as learning disabled, mildly retarded, and emotionally disturbed), who have higher rates of absenteeism than do regular students (Weitzman, 1985). Research on students at the high school level shows that absenteeism is also substantially higher in inner-city schools than in other high school settings.

Another work characteristic that has far-reaching implications for student success is punctuality. Arriving at school, completing assignments, and preparing for examinations on time could make a big difference in one's academic success. Unfortunately, no recent research has directly addressed the issue of student punctuality. Several decades ago, however, research (Dudycha, 1936; Turney, 1930) suggested that punctuality was related to measures of intelligence and grades earned. Dudycha reported that students who were consistently late in meeting their commitments may not recognize this pattern as a problem and, hence, had little motivation to improve their punctuality.

Student performance can also be undermined by poor organization of work

materials. Materials as basic as paper and pencil and the textbook are frequently forgotten by students; other materials are lost or misplaced. The disarray of students' desks or lockers hinders efficient retrieval of necessary materials, which, in turn, may impair performance. Kops and Belmont (1985) have shown that low-achieving students typically have difficulty in organizing their work on school tasks and in selecting the right materials for the task at hand. Although there is practically no research on student orderliness in the management of materials, Doherty and Conolly (1985) have suggested that student tidiness will likely elevate the teacher's rating of a student's performance.

On-task behavior has been one of the best researched correlates of academic performance (Cobb, 1972; Prater, Hogan, & Miller, 1992; Soli & Devine, 1976; Wheldall & Panagopoulou-Stamatelatou, 1991). On-task behaviors include such actions as looking at the teacher as he or she gives instructions, taking notes as the teacher is talking, asking questions of the teacher, responding to teacher questions, and remaining actively engaged in independent and small-group activities. Distraction is one of the biggest impediments to students sustaining attention to the task at hand: Unfortunately, many urban students stay on task no more than 20% of the time; and their patterns of work tend to be disjointed (Haberman, 1997).

How the student responds to supervision is another important facet of work habits. Does the student accept supervision when given and ask for supervision when needed? In his work preparing troubled youth for on-the-job success, Wells (1991) has emphasized the importance of teaching students how to interact respectfully with their supervisors, how to clarify confusing directions, how to say "yes" to supervisors, and

how to follow instructions. The classroom is replete with opportunities for students to learn these skills.

Another part of the formula for success in school and in the adult workplace is facility in teamwork. Industry is increasingly appointing employees to work on projects as a team (Erdman, 1992). Solitary effort and individual competition are viewed as less helpful than sharing ideas, collaborating on problem solving, and accomplishing team goals. Students who have these skills will likely accomplish more in school and prepare themselves well for the workplace. On the other hand, students who do not work well with others potentially put themselves at a major disadvantage both in school and in the adult workplace.

Although the research is limited, some studies suggest that work and study habits similar to those mentioned thus far are predictive of performance in school. Prior to 1960, much of the research on the relationship between study habits and achievement focused on college students (Traxler, 1954; Weinland, 1930). Since that time, researchers have also investigated the relationship between the work habits and achievement of primary students (Fuchs, Fuchs, & Phillips, 1994; Schroeder, 1978; Stipek, 1983); but little research has been done on this relationship for middle school students. Middle school is one of the most important periods for cultivating work habits and developing a career perspective that will result in wise selection of high school courses and expanded career opportunities as an adult (Sears, 1985). If students are not successful during the crucial middle school years, they may decide to drop out of high school (Schreiber, 1962).

Although a number of reasons have been advanced for academic failure, Elliott, Voss, and Wendling (1966) noted that 75% of academic failure at the high school level was the result of poor study and examination habits and that 75% of the students who dropped out of high school had the ability to do passing or even superior work. The need for study-skills training for all students at the secondary level was further substantiated by Cottier and Koehler (1978), whose survey of ninth grade social studies students clearly showed that study skills, especially note taking and awareness of patterns of organization, had not been acquired. It is likely that high schoolers' work habits have their origin in the years prior to high school, especially in the middle school years.

Before educators can delineate the role of work habits in middle school, they first need a reliable and valid means of assessing those work habits. A variety of approaches has been used to assess student work habits: direct observation, work products, self-report, and teacher ratings. Each approach has its advantages and disadvantages. For example, direct observation can be a highly valid and reliable method of determining how students are working; but it is very labor intensive. Work products provide conclusive evidence of how well students are working but do not provide evidence of what work habits account for good or bad work products. Self-report has the advantage of examining work habits from the student's phenomenological perspective, but the student's perspective can be greatly distorted. A student's perception of her actions may be inconsistent with an observer's perception of those actions. The major advantage of teacher ratings is that someone who is with the child several hours every day can synthesize myriad observations of the child into a coherent rating of the student's work

style. A disadvantage is that teachers may allow personal bias to color their overall impression of the student.

Of the various assessment approaches described above, teacher ratings are probably among the most cost effective. In a relatively small amount of time, teachers can condense impressions from hundreds of hours with the child into a systematic rating of the student's work habits. Edelbrock (1983) underscored some of the advantages of rating scales: "Rating scales have recognized advantages in that they are less costly and time-consuming than direct observation or psychological testing and typically yield more objective and reliable data than assessments based upon projective techniques and clinical interviews" (p. 293).

Statement of the Problem

Inasmuch as work habits appear to be an important predictor of students' achievement at other grade levels, educators need to investigate the role of work habits at the middle school level. An important contributor to that research would be the availability of a reliable and valid measure of student work habits for middle schoolers. Of the various approaches that might be used in assessing work habits, teacher ratings appear to be the most efficient. However, the psychometric adequacy of teacher ratings remains to be determined. Edelbrock (1983) stated that teachers are generally qualified to report upon academic performance and behaviors, including a student's difficulty following directions, inattentiveness, and inability to complete tasks. In contrast, other studies have questioned teachers' effectiveness in observing and judging student behavior in the classroom. For example, Jecker, Maccoby, Breitrose, and Rose (1964) assessed the

relative contribution of verbal and nonverbal cues to a teacher's accuracy in judging student comprehension. Teacher judgments of student comprehension based on nonverbal cues were less accurate than judgments based on verbal cues. Thus, the accuracy of teacher ratings of student work habits remains somewhat in doubt.

Research Questions

The purpose of this study is to evaluate the psychometric adequacy of a teacher rating scale of student work habits that has been used in prior research (Bush, 1999; Malone, 1998). This scale, the Student Work Habits Rating Form, was used as a criterion measure for two work-related constructs evaluated by both Bush and Malone: work ethic and student valuing of school. Both work ethic and student valuing were assessed by student self-report forms. Teacher ratings on the Student Work Habits Rating Form were moderately linked to the two work variables targeted by Bush and Malone. Although it was used as a criterion measure in these two studies, negligible analysis was done on the psychometric adequacy of the Student Work Habits Rating Form. Thus, the overall focus of the current study is to evaluate the psychometric adequacy of this scale. Specifically, this study will address the following research questions regarding the Student Work Habits Rating Form:

1. What is the adequacy of the specific items on the scale?
2. What is the internal consistency of the teachers' ratings?
3. What is the linkage between student work habits and their achievement ratings?
4. What clusters of items are embedded in the Student Work Habits Rating Form?

CHAPTER II

REVIEW OF LITERATURE

Definition of Work Habits

Although attempts have been made to differentiate between *work habits*, *study skills*, and *study habits*, definitions of these terms often overlap. Kitson (1916) categorized study habits as memory, concentration of attention, reasoning, and expression. Traxler (1954) divided study activities into work habits and study skills:

Work habits consist of the application of a combination of attitudes toward, and understanding of, general principles of study. Study skills are the specific abilities and techniques employed. Study skills are more closely related to general academic aptitude than are work habits. (p. 24)

O'Leary (1955) also attempted to differentiate between work habits and study skills. Work habits relate to "operating in such areas as self-evaluation, seeking advice so as to work effectively, general self-direction of one's activities and degree of perseverance" (p. 4). Study skills were defined as "the ability to use the index, making proper use of the dictionary, or other specified techniques which will result in an efficient use of study time" (p. 4).

Study skills, however, may include not only specific techniques but also a wide range of behaviors and attitudes, such as the use of time, mental strategies for storing and organizing information, motivational strategies, and concentration on tasks undertaken (Gadzella & Williamson, 1984). Specific study skills include listening, building

vocabulary, using a textbook, doing homework, solving problems, preparing for and taking tests, following directions, using resources, planning time, and taking notes (Davenport, Spiegel, Klein, Fitzpatrick, & Mueller, 1984).

The nature, utility, and modifiability of study skills have been delineated by Devine (1987):

They are those competencies associated with acquiring, recording, organizing, synthesizing, remembering, and using information and ideas found in school.

Many should be valuable in nonacademic settings, but all seem more or less indispensable for school success. All are teachable at all levels to all students.

(p. 5)

Scott (1973) applied a similar understanding to work habits in general: "Work habits refer to behavioral patterns and traits which contribute to the ability to attain maximum growth in the light of measured capacities" (p. 9).

Schroeder (1978) defined work habits as the student's "tendency to apply himself to a learning task by attending carefully to task explanation and assignment, maintaining attention to the task, assuming responsibility for task completion and accuracy, and generally displaying adequate self-direction in individual and group learning activities" (p. 2). Traxler (1954) reached a similar conclusion based on his review of several books and articles related to the development of study skills and study habits:

Work habits include keeping in good physical condition, planning a definite study schedule, forming regular habits in regard to time and place of study, avoiding distractions, getting clearly in mind just what is to be done, persistence,

independence of work, punctuality, sustained application, participation in group activities, systematic review, an attitude of weighing evidence and suspending judgment, and an attitude of constructive self-criticism. (p. 24)

Although the terms *work habits*, *study habits*, and *study skills* have some conceptual overlap, the term that best reflects the construct addressed in this study is *work habits*, referring to those actions that are evidence of concentration on the task at hand and that maximize one's chances for success in the classroom. Because some sources use the terms *study habits* and *study skills* to refer to student actions labeled as *work habits* in the current study, these sources will also be cited as background for this study.

Predictive Potential of Work Habits

Teachers may explain a student's success or failure by the student's ability, the amount or type of effort the student expends, the difficulty of the task, or chance factors (Cooper & Lowe, 1977). However, many studies have found a significant relationship between students' work habits and academic success. For example, in *Learning How to Study and Work Effectively* by Book (1926), the following statement showed that study habits and work habits laid the groundwork for achievement:

To learn how to work effectively means, therefore, the acquisition of a definite group of habits, and the problem of learning how to work effectively becomes one of acquiring the specific habits of action and thought which make one efficient or able to follow a definite and scientific procedure when confronted by his tasks. (p. 182)

Evidence of the correlation between teachers' perception of work habits and students' academic success has been provided by Tollefson, Melvin, and Thippavajjala (1990). They found that teachers cited lack of consistent effort as the most important reason for students' low achievement. They also found that teachers often assert that such student characteristics as lack of motivation and poor work habits are important for explaining students' low achievement. Teachers indicate that students' achievement can be affected by acquired characteristics, including poorly developed study skills, poor reading ability, and previous academic experiences, which include poor academic background and inadequate preparation (Tollefson et al.). Poor reading performance among children also relates to a pattern of poor work habits (Schroeder, 1978). On the other hand, poor work habits could be a result of poor performance. The frustration and discouragement that results from poor performance may contribute to incomplete assignments and poor attendance (Tollefson et al.).

The relationship between teachers' beliefs about the importance of good student work habits, teacher planning, and student achievement has been investigated by Fuchs, Fuchs, and Phillips (1994). They found that teachers with strong beliefs about the importance of good work habits took responsibility for planning and promoted student achievement in reading and math to a greater degree than did teachers with weak beliefs about the importance of good work habits (Fuchs et al.).

Moreover, Farkas, Grobe, Sheehan, and Shuan's (1990) analysis of teacher evaluations of the work habits, disruptiveness, and physical appearance of seventh- and eighth-grade students revealed that basic skills and student work habits, including

homework, class participation, effort, and organization, are the principal determinants of coursework mastery. Students' work habits are especially predictive of course grades. Similar to the Farkas et al. study, Takei, Johnson, and Clark (1998) discovered that a student's grade point average, as well as predicting the pupil's achievement test scores, is strongly related to teacher evaluations of the pupil's classroom behavior.

A variety of other studies have also examined the relationship of study skills and academic success. Gadzella and Williamson (1984) examined the relationships among study skills, self-concept, and academic achievement. Students' study skills and self-concepts were both related to academic success. Underachievers exhibited such undesirable characteristics as poor peer relationships, negative attitudes, social inadequacy, bad study habits, low self-concepts, and poor family relationships (Goodstein, 1979). O'Leary (1955) studied the relationship of work habits to over- and underachievement in 9th grade and concluded that a high correlation existed among work habits, achievement, mental age, and study skills. In addition, Zarb (1981) examined nonacademic variables in predicting academic success among 10th-grade students attending a commercial school. The nonacademic variables included study habits and self-concept in relation to family, peers, or attitudes toward school. The successful students were those who had developed good study habits and who perceived themselves to be academically successful.

Poor organizational skills appear to be a major contributor to underachievement (Diener, 1960; Greene, 1986; Raph, Goldberg, & Passow, 1966). Underachievers do not know how to take effective notes and outline, how to identify important information

when reading material from textbooks, or how to develop an effective system for remembering data (Greene). Many gifted underachievers lack the study skills necessary to excel in school. They may be getting adequate grades without considerable effort, but their work habits may lead to failure later in school and in the adult world. In "*Could Do Better*": *Why Children Underachieve and What to Do About It*, Mandel and Marcus (1995) concluded that these students do not know how to study, take notes, or review for tests. They note that good students recognize and cultivate a learning style that works best for them, but underachievers do not know how to tune in to their natural learning style and so fail to develop or use it.

In many studies, study skills and habits have been considered important contributors to success in high school and college. Keith, Reimers, Fehrmann, Pottebaum, and Aubey (1986) studied the relationship between total study time and academic achievement of high school seniors. They found that the amount of time spent on homework greatly influenced achievement. In contrast to Mandel and Marcus' findings (1995), Weigel and Weigel (1967) found that both successful and unsuccessful college students knew what actions constituted effective study skills; but only the successful college students used them. Levine (1976) employed a survey in his exploration of the relationship between study habits, attitudes, and GPA for Black and White male and female college students. That study found a positive correlation between study habits and grades for all White students and for Black females. A study by Locke (1977) found a significant correlation between completeness of college students' notes and academic success as measured by course grades.

Weinland (1930) demonstrated how work habits make a difference among college students. It was clear that successful students avoided distractions during study and that an ability to manage time is more important than good intentions. Successful students took complete notes, and they had enough interest in their work to recall and think about what they have heard and read. Organization of the material was furthered, and the memory was strengthened by this process.

Relationship of Grade Level to Work Habits

Elementary students have different work-habit deficiencies than do intermediate students. According to Schroeder (1978), the three most widely reported study deficiencies of elementary school children were failure to carefully read written directions and assignments, failure to tune out distractions while working, and failure to utilize time provided in class to complete learning activities. Work habits that undermine elementary children's achievement were inattention, careless work, and failure to complete assignments on time or at all. Schroeder contended that the development of good work habits was as important as reading instruction for elementary students with poor work habits.

An especially important study skill in middle school is note taking. Even though note-taking skills become particularly important around the fifth grade, many middle-grade students have negligible note-taking skills (Davenport et al., 1984). The student who entered junior high school not knowing how to take complete and accurate notes was not likely to succeed (Greene, 1986).

Interestingly, Evans and Tribble (1987) found that students in middle school scored significantly higher on self-reported study skills than did students in secondary school. Presumably, middle school students get more school support for study skills than do secondary school students. Teachers wanting to help students with the transition from elementary school classrooms to the greater diversity in the curriculum of the middle school setting may give more attention to middle school students' study habits (Evans & Tribble).

Strategies for Promoting Work Habits

Some poor work habits are already apparent in early childhood (Stipek, 1983), suggesting that parents are important contributors to the development of children's work habits (Kazanas & Wolff, 1972). Exhibiting good work habits in the classroom does not necessarily mean that students will do the same at home. Even though primary students were capable of planning study time at home, they needed guidance and support from their parents until they had learned to take responsibility for getting homework done (Graham & Robinson, 1984).

Georgiady and Romano (1994) suggested several steps that middle school students should take to increase their academic performance: attending school regularly, paying attention in class, and taking notes on information presented in class. Moreover, adolescents were better able to improve their study habits if they knew their own learning styles (Boveja, 1998). Contributing factors to good study habits included instruction in organizational skills and role models with effective work habits (Zarb, 1981). Because study skills are developmental in nature, children have to keep developing and refining

them from one grade to the next. It is critical to teach study skills in the early grades and reinforce them during later school years (Davenport et al., 1984). The following are some areas to consider when teaching students good work habits: classroom arrangement, study routines, encouragement, allowance for individual differences, clear instructions, and praise (Kazanas & Wolff, 1972).

Teachers should also help students budget their time and work efficiently, because failure to get work done on time is common in the classroom. Students often fail to finish the assignments they start, and they find it difficult to do schoolwork at home. To help students get off to a good start, teachers should make sure that students understand what must be done and when it is due (Graham & Robinson, 1984).

Teachers can help students in developing study skills by providing explicit instructions, demonstrations, and feedback about learning strategies (Rafoth & DeFabo, 1990). For example, writing lecture material on the blackboard dramatically increased its probability of appearing in students' notes. Stressing the importance of lecture materials not written on the board may also help improve student achievement (Locke, 1977).

Schroeder (1978) had other specific recommendations that teachers can implement to build a structure for developing good student work habits. The teacher's role is (a) to help the student understand what abilities will be gained as a result of task assignment, (b) to make consequences of good or poor performance clear at the beginning, (c) to guide students consistently in good task-performance skills, (d) to plan alternate methods of instruction if performance is not satisfactory, and (e) to help the students organize their approach to a task (Schroeder).

Teachers tend to interact differently with high and low achievers. Fagot (1973) found that preschool teachers tend to be less directive in their interaction with children with good work habits than children with poor work habits. Teachers gave fewer directions and offered less unrequested information to children with good work habits. Kerman (1979) observed that teachers are more supportive of high achievers and call on them more than they do low achievers. Teachers need to be aware of differences in their responses to high- and low-achieving students, making sure that low achievers receive the teacher attention and guidance necessary to develop good study skills.

Assessment of Work Habits

Researchers typically use students' self-reports, teacher ratings, or behavioral observations when assessing classroom behavior and its relationship to other variables (Ehman, 1970). An important method for study-skills measurement has been self-report (Evans & Tribble, 1987). Evans and Tribble (1987) examined study-skills measurement in grades 7 through 12, using a self-report Study Habits Questionnaire (SHQ) as a tool for individual and class diagnosis and evaluation of study skills. Scale reliability and validity of the Study Habits Questionnaire was considered satisfactory; Cronbach's alpha was .79, and the measure correlated significantly with student GPA. However, subscales for the SHQ were not supported by factor analysis.

College students took the CAI (computer-assisted instruction) Study Skills Test as a self-reporting instrument. Correlation between GPA and the total study-skills scores was significant ($r = .52, p < .05$; Gadzella & Williamson, 1984). Hirsch's study (1983)

on validity and reliability of the CAI Study Skills Test also found a significant correlation between the scores and GPA ($r = .56, p < .01$).

Teacher judgments of students' work habits have been an important means of assessing this variable (Farkas et al., 1990). Because teachers' ratings tend to correlate highly with student classroom behavior, teachers' ratings have been represented in some literature as a reliable and valid means of assessing student classroom behavior (Forness, Guthrie, & Hall, 1976; Meyers, Attwell, & Orpet, 1968). Anastasi and Urbina (1997) have found that rating scales are useful in assessing children and adolescents in educational settings. Natriello and Dornbusch (1983) analyzed objective data on teacher assessment of student behavior and concluded that teachers were influenced more by student behaviors than by their ascribed characteristics, such as sex and ethnicity.

The Classroom Standards Scale, a teacher rating scale, has nine items that describe competencies and skills related to students' work habits and classroom behavior. In Fuchs et al.'s study (1994), the elementary teachers rated each skill by using a Likert-type scale. The alpha coefficient for this scale was .91 for teachers, which indicates this scale's high reliability (Fuchs et al.).

Teacher ratings have been found generally to be consistent with student reporting. Study habits and attitudes of Native American students in grades 7 through 12 were investigated by Hurlburt, Gade, and Fuqua (1985). The Survey of Study Habits and Attitudes Form H (SSHA) was administered to students in this study. Classroom achievement and behavior were measured by teacher ratings of student academic achievement, cooperation, and work habits. This study focused on the relationship of the

SSHA scores to teacher rating of student academic achievement, cooperativeness, and work habits. Each of the teacher rating scales proved to correlate significantly with the SSHA scales (Hurlburt et al.).

Perhaps the most reliable and valid method of assessing student work habits is direct observation. This claim is based on the assumption that “the human observer is a good instrument of quantitative observation, that he is capable of some degree of precision and some degree of objectivity” (Guilford, 1954, p. 278). The direct observation of behavior can be used for all ages and for all types of classroom situations, especially if the observer is the teacher or parent who fits into the school environment (Anastasi & Urbina, 1997). The usefulness of teacher observation may vary, depending on the breadth of contact with the same students. For example, preschool children have one teacher who observes a child across several class periods, but high school teachers often have students for one class period during the day. Thus, the usefulness of teacher observation may decrease as a child advances in school and contact with any single teacher lessens (Kamphaus & Frick, 1996).

Several types of observational schemes have been used in assessing students’ in-class work activity: checklists, frequency counts, and interval recording. All require an operational definition of the work habits to be assessed and observers who can provide undivided attention to the targeted work habits. The most common measure of work activity is on-task behavior. Typically, increasing student attention to tasks and persistency with those tasks results in improved student performance (Cobb, 1972; Prater et al., 1992; Soli & Devine, 1976; Wheldall & Panagopoulou-Stamatelatou, 1991).

Direct observations of a student's behavior are important not only to provide an objective view of behavior but also to assess environmental contingencies of that student's actions (Kamphaus & Frick, 1996). Observations can be conducted by outside observers, but using specially trained observers can be very expensive. Teachers would be the best source of information on a student's attention problems and overactivity because they have the opportunity to observe the student in situations having high demands for sustained attention and inactivity. However, it is extremely difficult for the teacher to simultaneously teach and systematically observe a student's behavior. Although highly reliable and valid, direct observation is the most labor intensive of all the assessment approaches.

Development of Rating Scales

Teacher ratings represent a viable method of assessing student work habits, but they are not without limitations. Guilford (1954) demonstrated a number of errors that can occur when using rating scales. One of the most common sources of bias is the *halo effect*, defined as a tendency for raters' general impressions to bias their ratings of distinct aspects of the ratees. Another source of bias is *error of central tendency*. Some raters tend to avoid extreme categories and to place people in the middle of the scale. A third problem is *leniency error*. Some raters give scores that are consistently too high or too low (Guilford).

The Likert scale is one of the most popular formats for rating subject characteristics. It is used in instruments measuring opinions, beliefs, and attitudes (DeVellis, 1991). When using Likert-type scales, descriptive statements can be either

positive or negative (Fishbein & Ajzen, 1975; Nunnally, 1978). Nunnally recommended that an equal number of positive and negative statements should be selected. The purpose of using both positive and negative statements within the same scale is to avoid an acquiescence, affirmation, or agreement bias (DeVellis). In particular, this may also help to avoid a halo effect (Rao, 1973). In general, very mild statements may evoke too much agreement when used in Likert scales (DeVellis).

In developing a rating scale, one typically starts by generating a large pool of possible items thought to be consistent with the construct to be measured. Hinkin (1995) contended that generation of items might be the most important step in scale development. A response format, such as a Likert-type scale or an agree-disagree dichotomy, is then chosen. Items are selected from the pool, either on a conceptual basis by expert judges or on an empirical basis through factor analysis. Having experts review the item pool is important, because the review serves several purposes related to the content validity of the scale. Experts should rate how relevant they think each item is to what a scale developer intends to measure, evaluate the item's clarity and conciseness, and point out ways of tapping the phenomenon that the scale developer has failed to include (DeVellis, 1991). One of the methods of studying dimensions underlying relations among a set of items is factor analysis, which is based on the premise that the best approach to selecting items from the pool is to analyze the correlation matrix among all the items (Pedhazur & Schmelkin, 1991).

The sample of subjects should be a concern to the scale developer. Nunnally (1978) suggested that 300 people is an adequate number of subjects for the sample, but

scales have been successfully developed with smaller samples (DeVellis, 1991). The sample also should be representative of the population with which the rating scale will be used. Whatever the number and composition of the sample, the raters must have sufficient contact with students to judge them on the items included in the rating scale.

The reliability of the ratings is typically the first psychometric determination made. Internal consistency reliability is a function of how strongly the items correlate with one another and how many items the scale developer has in the scale (DeVellis, 1991). Although both internal consistency and test-retest reliability of the ratings are important to assess, internal consistency is probably the more fundamental characteristic to establish (Nunnally, 1978). Unless rater responses tend to cluster, one cannot claim to have a scale that measures a unitary construct or a set of related constructs.

Once reliability of the scale is secured, the test developer can focus on test validity. Two types of validity are particularly important in scale development: criterion related and construct related. Criterion-related validity is a matter of how strongly the rating scale is linked to other important variables to which it should logically be related. In the current study, for example, one could logically expect that teacher ratings of student work habits would be related to achievement measures. Construct-related validity is usually addressed through factor analysis, which identifies empirical clusters embedded within the scale. It is up to the scale developer to identify a meaningful conceptual label for each of these factors.

Finally, the contribution of each item to the total scale is assessed. This assessment may include a determination of the relationship between responses to the item

and the total scale scores, the discrimination potential of the item, the contribution of the item to the internal consistency of the scale, and the relationship between responses to the item and criterion measures. When evaluating the items, the following factors need to be considered: initial examination of items' performance, reverse scoring, item-scale correlations, item variances, item means, and coefficient alpha (DeVellis, 1991).

Conclusions of the Literature Review

Many studies have found a significant relationship between students' work habits and such academic measures as reading performance, course grades, and GPA at all levels of education. Elementary students have different work habit deficiencies than do middle school students. Work habits that often undermine elementary children's achievement are inattention, careless work, and failure to complete assignments on time or at all. Note-taking skills become particularly important around the fifth grade, but many middle-grade students have negligible note-taking skills. It is important to teach study skills in the early grades and reinforce them during later school years. The parents are important contributors to the development of their children's work habits, and teachers also can help students in developing study skills by providing explicit instructions, demonstrations, and feedback about learning strategies.

In general, researchers use students' self-reports, teacher ratings, or behavioral observations when assessing classroom behavior. Teacher ratings represent perhaps the most efficient way of assessing student work habits. The development of a rating scale typically involves the following steps: generating a representative item pool, having the initial item pool reviewed by experts, determining the format for participant responding,

administering the items to a large sample, and evaluating the psychometric adequacy of individual items. In particular, it is important to assess both the reliability and validity of the scale. In addition, the contribution of each item to the total scale should be assessed.

CHAPTER III

METHOD

This dissertation is part of a research program begun in 1996 on the assessment of the work values and work habits of middle school children. This research endeavor, known as the Middle School Survey project, has been supervised by Robert L. Williams, a professor in the Educational Psychology Department at the University of Tennessee, Knoxville. Several doctoral students, including the author of this dissertation, were involved in this research group from the beginning. These doctoral students assisted in developing the research questions, planning data collection, participating in data collection, and analyzing data. Two members of the group, Yvonne Malone and Vivian Bush, produced dissertations from their participation in this research. Malone's dissertation (1998) dealt with the psychometric development of the School-Valuing Scale, and Bush's dissertation (1999) presented the psychometric development of the Work-Ethic Scale. Both used the Student Work Habits Rating Form (the instrument targeted in this dissertation) as a criterion measure in their dissertations, but neither examined the psychometric adequacy of this scale. Therefore, this dissertation will provide a thorough evaluation of the psychometric properties of this scale. The data on which this dissertation is based were collected in the spring of 1997.

Subjects

The work habits of 238 sixth-, seventh-, and eighth-grade students from three U.S. middle schools were rated in this project. These students were selected by teachers from

a larger sample of 1,239 students who took the Work-Ethic Scale previously mentioned. All results from the data collection were first transmitted to the participating schools, which, in turn, gave written consent for our research group to use the results in written documents, such as journal articles and dissertations. The total research endeavor was approved as a Form A research activity.

The three schools were located in cities of varying sizes. School 1 ($n = 341$) was located in a northeastern city with a population of about 55,000. It was located in one of the economically poorer school districts in the state. It had the largest percentage of Hispanics students of any of the participating schools. School 2 ($n = 254$) was located in a southeastern city with a population of approximately 150,000. It was an urban school, with 100% of the students qualifying for free or reduced-price lunches. The principal ethnic group in School 2 was African American. School 3 ($n = 644$) was located in a suburban area of a southeastern city. Thirty-three percent of its students qualified for free or reduced-price lunches. The predominant ethnic group in School 3 was Caucasian, but the student body also included a substantial number of African Americans. The demographic composition of the work habits subsample ($n = 238$) used in the current study is presented in Table 1.

Description of Instruments

Social studies teachers ($n = 22$) in the three schools first rated all of their students on academic achievement ($N = 1,239$) and then rated the work habits of a subsample of students ($n = 238$) from each class. Only the achievement and work habits ratings of this

Table 1

Demographic Information for the Work Habits Subsample From the Three Schools

Demographic information	School 1	School 2	School 3	Total
Participants	71	55	112	238
Gender				
Males	37	23	59	119 (50%)
Females	34	32	53	119 (50%)
Ethnicity				
Caucasian	17	11	81	109 (46%)
African American	7	40	17	64 (27%)
Hispanic American	36	0	0	36 (15%)
Asian American	1	2	9	12 (5%)
Native American	0	1	1	2 (1%)
Other	10	1	4	15 (6%)
Grade				
Sixth	24	17	34	75 (31%)
Seventh	14	24	38	76 (32%)
Eighth	33	14	40	87 (37%)

subsample were used in the current study. Rating forms developed by the research group were used in rating both dimensions.

Achievement Rating Form

Teachers used their class records to rate students' academic achievement for the year as *poor*, *average*, or *outstanding*. In the framework of a traditional grading scale, a poor rating would be analogous to grades of D or F, an average rating would be comparable to C grades, and an outstanding rating would be analogous to grades of A or B. To preserve student anonymity, teachers used student numbers rather than names in supplying the requested data to the researchers. Scores on the achievement scale could range from 1 through 3 (i.e., from poor to outstanding; see Appendix A for a copy of the Achievement Rating Form).

Student Work Habits Rating Form

This scale was used by teachers to rate two students whom the teachers judged to have the best work habits in each class and two students judged to have the worst work habits in each class. The 20 items on this rating form described work-related behaviors, such as "Takes good care of books and other classroom materials" and "Makes good use of teacher feedback." Items for this rating form were generated by the original research group associated with the Middle School Survey project. Items reflected behaviors mentioned in the literature as contributing to student performance as well as student behaviors that group members perceived as constructive. Items that all of the original group members ($n = 5$) identified as important were selected for inclusion in the rating scale. Teachers used a 5-point Likert scale to indicate how similar each behavior was to

each student rated. Scores on this instrument could range from 20 to 100 (see Appendix B for a copy of the Student Work Habits Rating Form).

Procedures

To evaluate the psychometric efficacy of the teacher rating form of middle school students' work habits, the current study focused on the adequacy of developed items and the reliability, criterion-related validity, and construct-related validity of the scale.

Item Analysis

Item-total correlations (the 20 correlations between item scores and the total work habits form summative score excluding the item of interest) were estimated to conduct item analysis. Based on the known-groups approach, each item's *t*-value and effect size was calculated. In addition, Samejima's (1969) Graded Response Model was chosen for the item response theory (IRT) analysis of the Likert-scale items using the software program MULTILOG 6.0 (Thissen, 1991). MULTILOG implements a marginal maximum likelihood estimation (Bock & Aitkin, 1981) to estimate the item parameters.

Reliability

The author used Cronbach's α to estimate internal consistency. Also, marginal reliability was obtained from the IRT approach. Coefficient alpha, which is a measure of internal consistency derived from the average correlation among items and the number of items, is the first measure of reliability that should be obtained on a new instrument (Nunnally, 1978).

Validity

Criterion-related validity was assessed using both concurrent validity and a known-groups procedure. To estimate concurrent validity, a Pearson product moment correlation between the Work Habit Total score and the School Valuing Total score was computed. For the known-groups approach, the data from this study were analyzed using *t*-test, analysis of variance, and logistic regression analysis. Evidence for criterion-related validity was measured in terms of correlations with current performance on another valued measure (Gronlund, 1976).

Construct-related validity was assessed using both an exploratory factor analysis and a confirmatory factor analysis. In exploratory factor analysis, values in the diagonal of the correlation matrix were squared multiple correlations. Maximum likelihood estimation was used to estimate factor structures and loadings. Also, varimax and promax rotation methods were used. In this study, PRELIS 2 (Jöreskog & Sörbom, 1993a) was used to compute estimates of the asymptotic covariances. LISREL 8 (Jöreskog & Sörbom, 1993b) was used in confirmatory factor analysis (to test whether the one- or two-factor model was a good fit for the data under study).

CHAPTER IV

RESULTS

The frequency distribution, mean, and standard deviation (*SD*) for each item on the Student Work Habits Rating Form are in Table 2. Each item's frequency distribution shows a "U" shape that describes high frequency on the extremes and low frequency in the middle. The U-shaped distribution in this study is attributable to the exclusion of students judged to have average work habits. Thus, the sampling distribution does not satisfy the assumption of normality. As expected, the outstanding group's composite work-habits mean (93.54) was larger than the poor group's work-habits mean (46.36). All the differences between the two groups' item means were statistically significant; and effect sizes, computed by using the pooled *SD*, were large. The low group also had a larger *SD* on all items than the high group did. Ordinarily, effect size is based on the differences in group means divided by the standard deviation of the control group (Hair, Anderson, Tatham, & Black, 1998). Because the current study did not have a true control group, the pooled *SD* was used as the metric of comparison for the difference in the means of the high and low work-habits groups.

Item Analysis

The psychometric properties of the individual items on the Student Work Habits Rating Form were assessed through classical item analysis and Samejima's (1969) Graded Response Model (GRM) of item response theory (IRT). The latter approach was added to the item analysis because of limitations in classical measurement theory

Table 2

Frequency Distribution, Mean, SD, t-value, and Effect Size for Each Item

Item	Group ¹	Percentage of Responses					n	Mean	SD	t-value ²	Effect Size
		a lot like	somewhat like	neither like	somewhat unlike	very unlike					
1 *	Low	26.85	25.00	5.56	25.93	16.67	108	2.81	1.49	12.72	1.77
	High	0.77	1.54	0.77	14.62	82.31	130	4.76	0.62		
	All	12.61	12.18	2.94	19.75	52.52	238	3.87	1.47		
2 *	Low	25.00	25.93	11.11	27.78	10.19	108	2.72	1.37	14.36	1.99
	High	0.77	0.77	1.54	15.38	81.54	130	4.76	0.60		
	All	11.76	12.18	5.88	21.01	49.16	238	3.84	1.44		
3	Low	50.00	19.44	7.41	10.19	12.96	108	2.17	1.46	13.48	1.80
	High	3.08	7.69	3.85	11.54	73.85	130	4.45	1.08		
	All	24.37	13.03	5.46	10.92	46.22	238	3.42	1.70		
4 *	Low	12.96	18.52	10.19	26.85	31.48	108	3.45	1.43	9.88	1.39
	High	0.00	0.77	2.31	6.15	90.77	130	4.87	0.46		
	All	5.88	8.82	5.88	15.55	63.87	238	4.23	1.24		
5	Low	20.37	24.07	19.44	19.44	16.67	108	2.88	1.39	8.35	1.09
	High	4.62	8.46	9.23	9.23	68.46	130	4.29	1.21		
	All	11.76	15.55	13.87	13.87	44.96	238	3.65	1.47		
6 *	Low	43.52	28.70	10.19	12.96	4.63	108	2.07	1.22	22.01	3.07
	High	0.00	0.00	2.31	16.15	81.54	130	4.79	0.46		
	All	19.75	13.03	5.88	14.71	46.64	238	3.55	1.62		
7 *	Low	33.64	20.56	19.63	21.50	4.67	108	2.43	1.28	16.56	2.27
	High	1.54	1.54	0.77	15.38	80.77	130	4.72	0.71		
	All	16.03	10.13	9.28	18.14	46.41	238	3.69	1.52		
8 *	Low	37.96	22.22	12.96	19.44	7.41	108	2.36	1.36	17.03	2.37
	High	0.00	1.54	1.54	18.46	78.46	130	4.74	0.57		
	All	17.23	10.92	6.72	18.91	46.22	238	3.65	1.55		
9 *	Low	37.96	24.07	14.81	18.52	4.63	108	2.28	1.27	19.29	2.68
	High	0.00	1.54	0.77	13.85	83.85	130	4.80	0.52		
	All	17.23	11.76	7.14	15.97	47.90	238	3.65	1.57		
10 *	Low	16.67	17.59	17.59	21.30	26.85	108	3.24	1.45	11.15	1.56
	High	0.00	1.54	0.77	7.69	90.00	130	4.86	0.48		
	All	7.56	8.82	8.40	13.87	61.34	238	4.12	1.31		
11 *	Low	12.96	10.19	9.26	29.63	37.96	108	3.69	1.40	8.20	1.15
	High	0.00	2.31	0.00	6.92	90.77	130	4.86	0.51		
	All	5.88	5.88	4.20	17.23	66.81	238	4.33	1.17		
12 *	Low	37.04	35.19	6.48	12.04	9.26	108	2.21	1.31	19.20	2.67
	High	0.00	1.54	0.77	14.62	83.08	130	4.79	0.52		
	All	16.81	16.81	3.36	13.45	49.58	238	3.62	1.61		
13 *	Low	69.44	19.44	1.85	6.48	2.78	108	1.54	1.01	26.14	3.51
	High	0.77	0.00	7.69	26.92	64.62	130	4.55	0.71		
	All	31.93	8.82	5.04	17.65	36.55	238	3.19	1.73		
14 *	Low	70.37	13.89	5.56	8.33	1.85	108	1.57	1.04	27.19	3.72
	High	0.00	0.77	3.08	27.69	68.46	130	4.64	0.58		
	All	31.93	6.72	4.20	18.91	38.24	238	3.24	1.74		
15	Low	40.74	17.59	13.89	12.96	14.81	108	2.44	1.49	10.30	1.36
	High	4.62	11.54	3.85	10.00	70.00	130	4.29	1.24		
	All	21.01	14.29	8.40	11.34	44.96	238	3.45	1.65		
16 *	Low	54.63	22.22	9.26	8.33	5.56	108	1.88	1.21	22.00	3.03
	High	0.00	1.54	2.31	21.54	74.62	130	4.69	0.60		
	All	24.79	10.92	5.46	15.55	43.28	238	3.41	1.68		
17 *	Low	63.89	25.00	5.56	3.70	1.85	108	1.55	0.90	35.26	4.89
	High	0.00	0.00	0.00	17.69	82.31	130	4.82	0.38		
	All	28.99	11.34	2.52	11.34	45.80	238	3.34	1.77		
18 *	Low	51.85	26.85	12.04	7.41	1.85	108	1.81	1.04	28.97	4.04
	High	0.00	0.77	0.00	10.77	88.46	130	4.87	0.40		
	All	23.53	12.61	5.46	9.24	49.16	238	3.49	1.71		
19 *	Low	61.11	25.00	4.63	4.63	4.63	108	1.67	1.08	27.84	3.85
	High	0.00	0.77	1.54	15.38	82.31	130	4.79	0.49		
	All	27.73	11.76	2.94	10.50	47.06	238	3.37	1.76		
20	Low	65.74	19.44	5.56	4.63	4.63	108	1.63	1.09	16.51	2.15
	High	5.38	10.77	7.69	11.54	64.62	130	4.19	1.27		
	All	32.77	14.71	6.72	8.40	37.39	238	3.03	1.75		

Note: * Items are reverse scored.

¹ Students are grouped according to the teacher's judgment of the student's overall work habits.² All t-values are statistically significant at $p < .0001$.

(Hambleton, Swaminathan, & Rogers, 1991). First, both examinee and test characteristics are interrelated and cannot be separated in classical item analysis. Thus, test and item characteristics change as the examinee context changes (group dependence); and examinee characteristics change as the test context changes (test dependence). In other words, group dependence means that item characteristics depend on the group of examinees; and test dependence means that examinee scores depend on the particular set of items administered. Second, two sources of dissatisfaction with classical test theory are the definition of reliability based on parallel forms (impossible or difficult in practice) and the assumption of equal errors of measurement for all examinees (implausible). Third, classical test theory is test-oriented rather than item-oriented. Thus, classical test theory does not provide the probability of an examinee's answering a given item correctly. Such information is necessary, for example, to design a test intended to discriminate well among scholarship candidates. In contrast, IRT provides the following psychometric advantages:

(a) item characteristics that are not group-dependent, (b) scores describing examinee proficiency that are not test-dependent, (c) a model that is expressed at the item level rather than at the test level, (d) a model that does not require strictly parallel tests for assessing reliability, and (e) a model that provides a measure of precision for each ability score. (Hambleton et al., 1991, p. 5)

In addition, information from IRT is especially useful for creating short, equated, and multiple forms of a scale. This possibility is due to the separation of item parameters from examinee (subject) parameters in IRT (Uttaro & Lehman, 1999). The item

parameters (i.e., slope and difficulty) are not affected by the assumed ability levels of the sample. For this reason, IRT is characterized as a sample-invariant procedure.

In short, the major characteristic of IRT is that it provides the chance to examine the relationship between the individual's response to a set of items and the latent trait of interest (e.g., work habits; Kirisci, Moss, & Tarter, 1996). Latent trait refers to the construct assumed to be embedded in a set of items. The central advantage of the GRM is that more information about an individual's ability or attitude level can be obtained with this model than with the classical item analysis (Koch, 1983). In particular, GRM is useful when the response categories are ordered, as in a Likert-type scale (Samejima, 1969, 1997).

According to the results of classical item analysis in this study, the range of item-total correlations between item score and the total score excluding the item of interest (.93 to .53) indicates that all items were highly or moderately correlated with the Student Work Habits Rating Form (see Table 3). According to the *t*-test and effect-size results, poor- and outstanding-achievement groups showed significant differences on all items.

The Inter-Item Pearson Product Moment Correlation Matrix (see Appendix C) indicates that some items were very highly correlated, thus producing problems for factor analysis (Tabachnick & Fidell, 1996) and for IRT analysis (Lynch, 1997). For example, maximum likelihood estimation, a popular method of common factor analysis, yields unstable estimates when the items are very highly correlated (Nunnally & Bernstein, 1994). An example of this instability is the obtaining of correlation coefficients greater than 1.00 in the regenerated correlation matrix on which the factor analysis is based.

Table 3

Classical Item Statistics and Item Estimates From Samejima's Graded Response Model of IRT

Item	Classical Item Statistics						Graded Response Model			
	Item-Total Correlation ^a	Rank ^b	<i>t</i> -test ^c	Rank	Effect Size ^c	Rank	Slope (S.E.) ^d	Rank	Information ^e	Rank
1	.81	13	13.32	15	1.93	15	2.67 (.35)	12	11.38	12
2	.83	11	16.15	13	2.34	13	2.74 (.33)	11	12.63	11
3	.73	16	15.37	14	2.17	14	2.00 (.22)	16	6.79	16
4	.67	17	10.10	18	1.47	18	1.79 (.29)	17	5.77	17
5	.56	19	8.34	19	1.15	19	1.24 (.19)	19	3.66	19
6	.87	7	24.63	6	3.56	6	3.09 (.35)	8	13.94	8
7	.84	10	16.75	12	2.40	11	3.24 (.41)	7	15.76	7
8	.82	12	17.28	10	2.48	10	2.42 (.26)	14	10.18	13
9	.88	6	21.07	8	3.05	8	3.39 (.42)	6	16.10	6
10	.73	15	10.88	17	1.58	17	2.26 (.32)	15	8.71	15
11	.53	20	7.54	20	1.09	20	1.23 (.23)	20	2.87	20
12	.86	9	20.03	9	2.88	9	2.90 (.34)	10	12.79	9
13	.87	8	26.24	5	3.60	5	2.98 (.34)	9	12.74	10
14	.92	4	30.76	2	4.35	3	4.11 (.43)	4	19.60	4
15	.67	18	11.37	16	1.59	16	1.58 (.19)	18	5.03	18
16	.90	5	21.70	7	3.10	7	3.52 (.41)	5	16.47	5
17	.93	2	36.49	1	5.17	1	5.14 (.71)	2	25.92	2
18	.93	1	30.43	4	4.34	4	5.51 (.83)	1	29.50	1
19	.93	3	30.76	3	4.43	2	4.39 (.55)	3	20.61	3
20	.78	14	17.18	11	2.36	12	2.45 (.26)	13	9.14	14
Cronbach's α (internal consistency) = .98 and Marginal Reliability ^f = .96										

Note: ^aCorrelation between item score and the total score excluding the item of interest.

^bValues were ranked (descending order) by their sizes.

^cStudents were grouped according to academic achievement (poor, average, and outstanding).

For the *t*-test and the effect size computation, poor and outstanding groups were used.

^dS.E. = Standard error for slope.

^eInformation is generally higher when the slope parameter is high.

^fMarginal reliability from the IRT results.

The results from GRM of IRT showed that estimates of item-slope parameters were closely related to the item-total correlation. The item-discrimination parameter (a) indicates the slope of the curve (Nunnally & Bernstein, 1994). The higher the value of a , the steeper the slope (Anastasi & Urbina, 1997). Items with steeper slopes are more useful for discriminating among the ability levels of the examinees than are items with flatter slopes. The typical range of a is 0 to 2 (Hambleton et al., 1991).

Important features of IRT are its treatment of reliability and error of measurement through item-information function, which is computed for each item (Anastasi & Urbina, 1997). Item-information function is a powerful method to describe items and tests, to select test items, and to compare tests' shortcomings (Hambleton et al., 1991). The information function indicates the pattern of precision over various levels of θ (or the trait being assessed). This information function mitigates the limitation of only one reliability estimate generated via the classical-test theory. The higher the value of the information function in IRT, the more precisely the item is being measured. In Appendix D, item-information functions show the measurement efficiency or precision of the item at different levels of work habits. For example, Item 18 shows the highest information at $\theta = 0.0$, and Item 9 shows the highest information at $\theta = -0.5$.

The Student Work Habits Rating Form is very useful to measure latent trait levels from -0.5 to 0.5, based on the results that test information at $\theta = -0.5$ is 56.25, test information at $\theta = 0.0$ is 59.74, and test information at $\theta = 0.5$ is 52.67. The test-information function is the sum of the individual item-information functions. The overall test-information function indicates where the test discriminates best and most poorly so

that items can be selected for test subjects (Nunnally & Bernstein, 1994). Although the sample used in this study included only high and low groups, the overall information test function indicates that the Student Work Habits Rating Form discriminates best for subjects obtaining ratings toward the middle part of the distribution.

The two-parameter logistic model was applied to examine both difficulty and discrimination. Individual-item information is summed over ability or Work-Habit (θ) levels (-2.0 to +2.0, categories by .5; see Table 3). When the ability values of a group are transformed so that their mean is 0 and their standard deviation is 1, the values of difficulty vary typically from about -2.0 to +2.0 (Hambleton et al., 1991). Additional information regarding GRM values is in Appendix D. Values near -2.0 correspond to easy items, and values near +2.0 correspond to difficult items. There are four thresholds (difficulty) in the Student Work Habits Rating Form. The b values in Appendix D reflect these four thresholds. In general, a threshold is one less than the number of ordered responses in a particular model. Each threshold specifies the point on the θ at which an examinee has a 50% chance of responding in some higher category than the one to which the threshold belongs. Appendix D indicates that progressing from neutral to positive levels on the Likert-type scale is the major point of transition from easy to difficult ratings for a considerable number of the items. Items 3 and 13 through 20 represent the most difficult items in this respect. An easy item is one in which students with poor work habits can achieve a positive rating, whereas a difficult item is one in which only students with good work habits are likely to receive high ratings. For example, Item 11 (practices

good attendance) is an extremely easy item, whereas Item 20 (is easily distracted) is a far more difficult item.

The correlation between slope parameters and item-total correlations was .881, and the correlation between item information and item-total correlations was .869. As mentioned previously, the typical range of slope parameters (item discrimination) is 0 to 2. Although high values for slope are desirable, it is unusual to obtain values larger than 2 for slope parameters. However, in this study, 15 items (out of 20) have values greater than 2. The possible explanation is the U-shaped item-frequency distributions.

For the purpose of evaluating items, the classical approach and the IRT approach gave similar conclusions. For example, Items 14, 16, 17, 18, and 19 appear to be good items; whereas Items 4, 5, 11, and 15 seem to be less-than-good items within the Student Work Habits Rating Form (see Table 3). This conclusion is based on item-total correlations and effect sizes in Classical Item Statistics and the slopes and information function of the Graded Response Model. An examination of Appendix B shows that good items seem to be directly related to the activity of work, whereas the less-than-good items appear to reflect behaviors prior to or outside of work.

Reliability

Reliability of the Student Work Habits Rating Form was estimated using Cronbach's α . An obtained coefficient was .98, and marginal reliability based on IRT was .96 (see Table 3). The marginal reliability describes an average of the reliability over the various levels of θ (Lynch, 1997). The obtained α substantially exceeds the minimal level of .70 suggested by Nunnally (1978) for research instruments.

Validity Analysis

Using the known-groups approach based on the three achievement groups, the author also examined the criterion-related validity for the Student Work Habits Rating Form. Table 4 shows that the work habit (WH) means for the three achievement groups were 44.92 for the poor group, 78.16 for the average group, and 94.45 for the outstanding group. Tukey's Studentized Range Test revealed that the differences between means of all three groups were statistically significant at $\alpha = .05$. In addition, the logistic regression analysis indicated that, using the WH scores, 98.1% of members in the poor or outstanding achievement groups were predicted correctly. Although this study had three achievement groups, the sample size of the average group (25 students) was relatively small compared to the other groups (the poor group had 99 students and the outstanding group had 114). Thus, only the poor and outstanding groups were used in the logistic regression analysis. Table 5 shows the results of exploratory factor analysis. The results can be used for two related purposes: (a) to determine construct validity and (b) to determine unidimensionality (also related to local independence) for IRT and scale development. A common assumption of IRT is that only one ability (a dominant factor influencing test performance) or only one latent trait is measured by a set of items in a test. In other words, the latent trait (e.g., work habits) specified in the model is assumed to be the only factor influencing examinees' response to test items. This unidimensionality is related to "local independence." Local independence means that item responses are independent when the abilities influencing test performance are held constant (Hambleton et al., 1991).

Table 4

*Results for Criterion-Related Validity From the Known-Groups Approach***Descriptive Statistics for Work Habits by Achievement Groups**

	Achievement Group		
	Poor	Average	Outstanding
Mean	44.92	78.16	94.45
SD	12.84	17.41	7.94
<i>n</i>	99	25	114
Minimum	24	36	54
Maximum	77	100	100

Analysis of Variance for Work Habits by Achievement Groups

Source	DF	Sum of Squares	Mean Squares	F Value
Model	2	130990.8	65495.4	503.5
Error	235	30566.9	130.1	(<i>p</i> < .0001)
Corrected Total	237	161557.7		

Note: $R^2 = .811$.

Tukey's Studentized Range Test revealed that the mean differences among the three groups were statistically significant at $\alpha = .05$.

Classification Results From Logistic Regression Analysis

Actual Achievement Group	Predicted Achievement Group		Total
	Poor	Outstanding	
Poor	97	2	99
Outstanding	2	112	114
Total	99	114	213

Correct Classification % = 98.1

Note: In this logistic regression analysis, only the poor and the outstanding achievement groups were used.

Table 5

Results of Exploratory Factor Analysis: One- and Two-Factor Models

One-Factor Model			Orthogonal Two-Factor Model			Oblique Two-Factor Model		
Factor 1			Factor 1			Factor 1		
Item	Pattern	h^2	Pattern	Pattern	h^2	Pattern	Structure	Structure
1	0.789	0.623	0.428	0.790	0.808	0.157	0.714	0.892
2	0.806	0.649	0.459	0.770	0.803	0.210	0.734	0.884
3	0.724	0.524	0.524	0.516	0.540	0.428	0.689	0.671
4	0.651	0.424	0.334	0.676	0.568	0.092	0.581	0.751
5	0.535	0.287	0.322	0.477	0.332	0.180	0.489	0.562
6	0.884	0.781	0.738	0.485	0.779	0.729	0.872	0.720
7	0.837	0.700	0.592	0.621	0.736	0.464	0.795	0.794
8	0.830	0.689	0.689	0.462	0.688	0.676	0.818	0.681
9	0.888	0.789	0.691	0.562	0.794	0.627	0.861	0.775
10	0.709	0.502	0.362	0.740	0.678	0.096	0.633	0.752
11	0.531	0.282	0.408	0.340	0.283	0.366	0.512	0.205
12	0.882	0.779	0.777	0.423	0.783	0.813	0.882	0.097
13	0.908	0.825	0.832	0.391	0.845	0.903	0.919	0.023
14	0.940	0.883	0.802	0.491	0.883	0.810	0.932	0.171
15	0.645	0.417	0.403	0.557	0.472	0.246	0.596	0.490
16	0.928	0.862	0.797	0.480	0.865	0.810	0.924	0.159
17	0.960	0.922	0.868	0.434	0.943	0.929	0.970	0.058
18	0.953	0.908	0.797	0.522	0.907	0.788	0.941	0.214
19	0.955	0.913	0.850	0.449	0.924	0.896	0.959	0.089
20	0.785	0.616	0.651	0.436	0.614	0.639	0.773	0.187
Eigenvalue	13.372		8.301	5.944	14.245			
Total Variance	66.9%		41.5%	29.7%	71.2%			
Common Variance	66.9%		58.3%	41.7%				
Inter-Factor Correlation from the Two-Factor Oblique Model = .714								
Significance Test								
H0: 1 factor is sufficient			$\chi^2 (170) = 687.4, p = .0001$			Tucker and Lewis's Reliability Coefficient = .90		
H0: 2 factors are sufficient			$\chi^2 (151) = 392.1, p = .0001$			Tucker and Lewis's Reliability Coefficient = .95		

Note: h^2 is the square of the factor loadings, indicating the amount of the variance explained in the item-factor relationship.

The one-factor model explained 66.9% of total variance and the two-factor model explained 71.2% (see Table 5). So, one more factor explained only 4.3% more of the variance. Factor loadings represent “the correlation between each original item and the factor from which the path emanated” (DeVellis, 1991, p. 95). The range of the factor pattern (factor loadings) for the primary factor was from .531 to .960. All item loadings (patterns) in the one-factor model were greater than .5, which was enough to satisfy the suggestion that items with factor loadings should be equal to or greater than .3 (Nunnally, 1978) or .4 (Tabachnick & Fidell, 1996). According to the scree test (see Figure 1) and the Tucker and Lewis Reliability index (.90), we can accept the one-factor model as an adequate fit for the data.

Table 6 presents the results of a confirmatory factor analysis (CFA). The purpose of the CFA in the present study was to assess whether the factor structure derived in the exploratory analysis provided a good fit for the data. The one-factor model was confirmed in this analysis. The χ^2 goodness-of-fit test was used as a test of fit between the various factor models and the data on which the factor analyses were based. Lower χ^2 values indicate a better fit between a model under study and the data than do higher χ^2 values. In addition to the χ^2 test, RMSEA, GFI, AGFI, NFI, and CFI were calculated to obtain goodness of fit indexes. For the RMSEA index, values below .05 suggest a good fit; whereas, for GFI, AGFI, and CFI, values greater than .90 typically indicate a good fit. All of the model-fit indexes of the one-factor model and the two-factor oblique model yielded better values than for the orthogonal two-factor model. Both the one-factor model and the oblique two-factor model were statistically acceptable models because χ^2 s'

Figure 1

Scree Plot of Eigenvalues From Maximum Likelihood Estimation

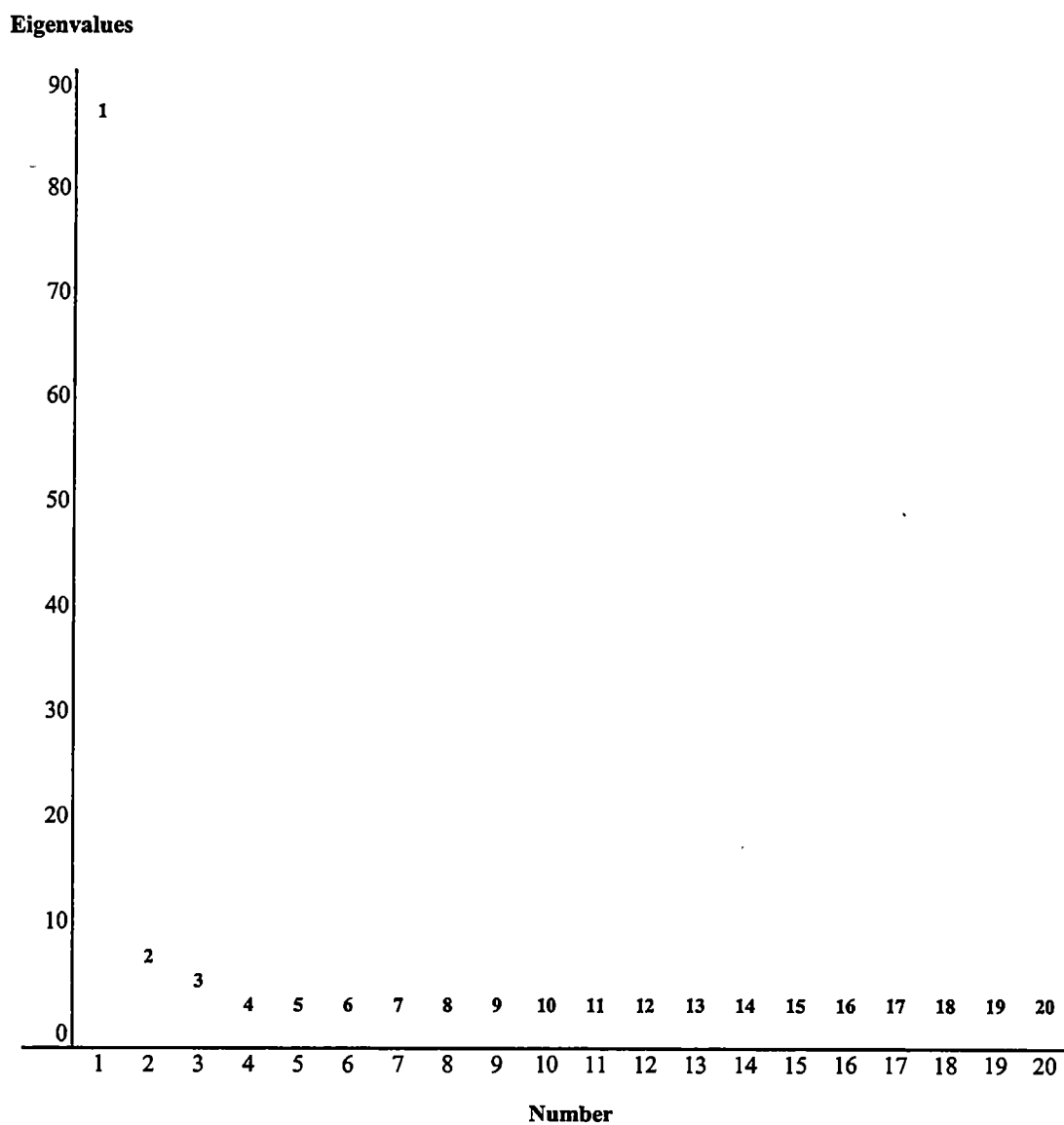


Table 6

Confirmatory Factor Analysis Results: One- and Two-Factor Models

Item	One-Factor Model	Orthogonal Two-Factor Model		Oblique Two-Factor Model	
		Factor 1	Factor 2	Factor 1	Factor 2
1	0.964		1.000*		1.000*
2	0.959		0.871		0.995
3	0.885	0.331		0.885	
4	0.865		0.464		0.900
5	0.737		-0.100		0.765
6	0.953	0.771		0.954	
7	0.948	0.483		0.949	
8	0.920	0.907		0.923	
9	0.962	0.858		0.964	
10	0.923		0.434		0.959
11	0.780	0.653		0.777	
12	0.951	0.888		0.950	
13	0.967	0.951		0.968	
14	0.989	0.965		0.990	
15	0.839		0.063		0.878
16	0.974	0.939		0.973	
17	1.000*	1.000*		1.000*	
18	0.993	0.958		0.993	
19	0.995	0.989		0.997	
20	0.923	0.650		0.922	
χ^2 Test	$\chi^2 (170) = 63.3, p = 1.00$	$\chi^2 (170) = 3449.4, p = 0.00$		$\chi^2 (169) = 57.5, p = 1.00$	
RMSEA	N/A **	.285, $p = 0.00$		N/A	
GFI	N/A	.923		N/A	
AGFI	N/A	.905		N/A	
NFI	N/A	.918		N/A	
CFI	N/A	.922		N/A	
Inter-Factor Correlation = .969 (Standardized Solution)					

RMSEA = Root Mean Square Error of Approximation; GFI = Goodness of Fit Index;

AGFI = Adjusted Goodness of Fit Index; NFI = Normed Fit Index; CFI = Comparative Fit Index.

*For the scale of factors, 1.0 was assigned. This is a typical approach for the scaling of factors.

**Because of perfect fit between the model and the data ($p = 1.00$), these values were not computed.

p values were 1.00. However, the inter-factor correlation in the oblique two-factor model was .969, which suggests that the factors represent highly similar constructs. In addition, the items of Factor 2 do not conceptually measure something distinct from Factor 1. Thus, the above results point to the one-factor model as a best model to explain the data under study. The results support not only the one-factor assumption of IRT but also the construct validity of the Student Work Habits Rating Form.

CHAPTER V

DISCUSSION

This chapter first summarizes the findings for the four research questions, then presents the current psychometric status of the Student Work Habits Rating Form, and lastly delineates areas for further development of the rating form. The data analysis revealed that this rating form would currently be appropriate as a research instrument for assessing student work habits but that further development is needed before using the rating scale on a clinical basis to assess the work habits of individual students. The latter is especially the case if decisions about interventions are to be made on the basis of a student's overall work habits or pattern of work habits.

Findings for Research Questions

The first research question focused on the adequacy of the specific items on the scale. The classical item and IRT approaches gave similar conclusions regarding the psychometric adequacy of specific items. For example, Items 14, 16, 17, 18, and 19 appear to be good items, whereas Items 4, 5, 11, and 15 seem to be less-than-good items within the Student Work Habits Rating Form. Good items appear to be directly related to the activity of work, whereas the less-than-good items appear to reflect behaviors prior to or outside of work.

The second research question dealt with the internal consistency of the teachers' ratings. Internal consistency of the Student Work Habits Rating Form was .98, which was very high.

The third research question related to the linkage between student work habits and their achievement ratings. The work-habits ratings for the three achievement groups judged as poor, average, or outstanding were significantly different. Also, the Student Work Habits Rating Form could classify correctly students in the poor and outstanding achievement groups almost 100% of the time.

The final research question dealt with the clusters of items that are embedded in the Student Work Habits Rating Form. The construct validity of the Student Work Habits Rating Form was assessed through a series of factor analyses, which pointed to a general factor of work habits. The one-factor model explained 66.9% of total variance, and the two-factor model explained 71.2%. Thus, the second factor appears to add little to the explanation of the variance.

Psychometric Status of Student Work Habits Rating Form

A primary strength of the Student Work Habits Rating Form is its internal consistency. A psychometric instrument cannot qualify as an appropriate instrument for research use and especially for clinical use without adequate internal consistency. This rating form greatly exceeds the .70 minimum traditionally required for research use. The internal consistency in the high .90s for this instrument exceeds that reported for most recognized instruments used in the assessment of student work values or work habits. For example, Cronbach's α of the Study Habits Questionnaire used by Evans and Tribble's (1987) study was .79, and the alpha coefficient for the Classroom Standards Scale used by Fuchs et al. (1994) was .91. Thus, the Student Work Habits Rating Form

offers considerable promise as a reliable means of assessing the work habits of middle school students.

Another important issue in judging the psychometric adequacy of any scale is its criterion-related validity. The major outcome measure used in judging the validity of the Student Work Habits Rating Form was teacher ratings of student achievement. Using student grades in social studies classes as their frame of reference for rating student achievement, the teachers judged each student's academic performance as poor, average, or outstanding. The work-habits ratings for the three achievement groups were substantially and significantly different. Plus, a logistic regression analysis showed that the work-habits ratings could correctly classify students in the poor and outstanding achievement groups virtually 100% of the time. Although only one outcome measure was used in the assessment of the criterion-related validity of the Student Work Habits Rating Form, the rating scale appears to have substantial potential as a predictor of student performance.

The construct validity of the Student Work Habits Rating Form was assessed through a series of factor analyses, culminating in Confirmatory Factor Analysis. This series of analyses points to a general factor of work habits rather than a set of primary and secondary factors. The finding of a general factor might have been expected from the high internal consistency of the scale. Thus, this rating form provides a coherent picture of student work habits rather than differential ratings of clusters of student work habits. As such, it provides an overall prediction of student success rather than a diagnostic profile of areas where work habits are strong or weak.

Although the Student Work Habits Rating Form showed considerable potential as a way to assess student work habits, all items on the rating form were not useful to an equal degree. Both classical item analysis and item response theory were used in evaluating the utility of the items. Classical item statistics typically address the proportion of the sample likely to get a favorable rating on an item, the correlation between each item and the total score on the instrument minus that item, and the relation of each item to criterion measures. Item response theory offers some psychometric advantages over classical item analysis (see the Results Chapter). A primary advantage is that conclusions regarding the adequacy of items are not as dependent on the specific sample used as is the case with classical item statistics. An additional advantage is that item response theory provides more information regarding the likelihood that a person who is strong or weak in work habits will receive a favorable rating on a particular item.

Several conclusions can be drawn from the various item analyses: (a) some items are more difficult than others with respect to receiving a favorable rating, (b) progression from low to high on the Likert-type scale is more difficult at the threshold from neutral to positive ratings than at the low end of the scale, and (c) most items discriminate best for students obtaining ratings toward the middle part of the work-habits distribution. With respect to the difficulty level of items, Item 11 (practices good attendance) proved to be the easiest item for obtaining a favorable rating and Item 20 (is easily distracted) the most difficult on which to achieve a favorable rating.

In the student's transitioning from neutral to positive ratings on the Likert-type rating scale, Items 3 and 13 through 20 appeared most difficult. These items reflect a

variety of positive characteristics (expresses enthusiasm about school work, takes initiative to keep busy, makes good use of teacher feedback, does good quality work, follows instructions well, and works at a steady pace) and negative characteristics (needs much supervision, pesters other students, and is easily distracted). These items seem to embody the possibility of exceptional characteristics for middle school students (i.e., attributes that one might not ordinarily expect from middle school students).

On the other hand, items on which it is easier to progress from a neutral to a positive rating (Items 1, 2, and 4 through 12) appear to reflect more of what might be expected from middle school students. Whatever theme might be subsumed in the latter set of items, students who are generally low on positive work habits are more likely to get favorable ratings on these items than on those items evidencing more difficulty in going from neutral to positive ratings.

The classical item and IRT approaches point to similar conclusions regarding the psychometric adequacy of specific items. The items that appear to have the strongest values in terms of slope, difficulty, and overall information function are the following: takes initiative to keep busy, makes good use of teacher feedback, does good quality work, follows instructions well, and works at a steady pace. In contrast, the items that appear less-than-good include arrives at class on time, easily gets upset, practices good attendance, and pesters other students. The items that best discriminate between good and poor workers relate to features of work itself, whereas those that discriminate less well relate to actions preparatory to or outside of work. Nonetheless, the weaker psychometric items have immense practical implications. Poor attendance and tardiness,

for example, are among the most important problem areas in the world of work. Likewise, emotional stability (easily gets upset) and good social relationships (pesters other students) can be fundamentally important to a worker's success. Thus, the practical utility of items should be considered in addition to their psychometric features in making decisions regarding the value of items.

As a comprehensive rating scale, the Student Work Habits Rating Form has strong internal consistency and moderately good criterion-related validity. It provides a coherent assessment of student work habits. Some items appear psychometrically stronger than others, but all items (including the psychometrically weaker items) appear to have practical utility.

Areas for Further Development

Despite the psychometric and practical potential of the Student Work Habits Rating Form, several features of this study limit the generalizability of the results and the clinical utility of the rating scale. Those limitations relate to the specific habits identified in the rating form, the nature of the student sample rated, the diversity and cross-validation of teacher raters, the absence of a test-retest reliability assessment for the rating scale, and the absence of diverse criterion measures for validating the scale.

First, the current Student Work Habits Rating Form has an imbalance of positive compared to negative items. Only 4 of the 20 work habits rated were stated negatively. Some researchers (Schriesheim & Hill, 1981) suggest that negatively stated items are more difficult to understand and thus detract from the accuracy of responding. On the other hand, a scale consisting primarily of positive items creates a climate conducive to

acquiescent responding (i.e., agreeing with items irrespective of their content; Kline, 1986). Because many raters have a tendency to agree with items, including mainly positive items in a rating scale makes it difficult to separate an acquiescent response tendency from genuinely favorable responses. Conversely, having a balance of positive and negative items permits a cleaner assessment of the acquiescent response tendency (Messick, 1962).

Some recent studies (Bush, 1999; Malone, 1998) indicate that negative items on a self-report measure also may elicit more accurate responses than do positive items. The proportion of negative items in the Student Work Habits Rating Form could be increased by reversing the wording of some positively stated items (“shows little interest in school work” rather than “expresses enthusiasm about school work”), adding a companion negative item for each positive item (“does poor quality work” as well as “does good quality work”), and including negative items not currently reflected in the scale (e.g., “jumps to conclusions,” “wastes time”).

The review of the work-habits literature uncovered a number of student work habits not included in the current rating scale: wastes time (Schroeder, 1978), gets work done on time (Graham & Robinson, 1984; Schroeder), is well organized (Farkas et al., 1990; Kops & Belmont 1985; Zarb, 1981), persists with work (O’Leary, 1955; Traxler, 1954), accurately evaluates own work (O’Leary; Traxler), plans study activities (Traxler), jumps to conclusions (Traxler), completes homework (Schroeder), works carelessly (Schroeder), and budgets time well (Graham & Robinson). In revising the item structure of the work-habits rating form, other work habits targeted in the literature should be

included. Perhaps some of the omitted work habits could replace the psychometrically weaker items in the current scale, particularly those more tangential to work. Certainly, the content validity of the Student Work Habits Rating Form could be challenged on the basis of its omission of important work habits reflected in the literature.

Second, the generalizability of the findings from the current study is limited by the student sample used. Only those students first identified by their teachers as having the worst and best work habits in their classes were rated. Such an entry-level distinction makes for an extremely bipolar distribution. It also creates a condition conducive to extreme ratings of work habits. Indeed, the ratings of students on the specific work habits were highly consistent with the teachers' initial discrimination between best and worst workers. The overall scores of those judged to be best workers on the Student Work Habits Rating Form were nearly double those judged to be the worst workers.

Preferable to asking teachers initially to identify their poorest and best workers would be to have them rate the work habits of all students in their class without prior classification of those students. This procedure would minimize the likelihood of the teacher's initial classification of a student dictating the way the teacher rates that student's specific work habits. Furthermore, a rating of all students would provide a better test of the rating form's potential to discriminate at all levels of student work habits. Plus, most of the statistical analyses typically done in scale development are more appropriate for a normal distribution than a bipolar distribution.

The sample to be used in the further development of the Student Work Habits Rating Form not only needs to include the middle portion of the work-habits distribution,

but it also needs to include a more representative sample of ethnic groups. Although Caucasian students were overly represented in the current sample, the study also had a reasonable sampling of Hispanic and African American students. Underrepresented were Native American and Asian American students. A more representative sample of various ethnic groups would permit comparison of teacher perceptions of student work habits for different ethnic groups.

One of the most significant limitations of the current study is the small number of teachers who did the ratings of student work habits. Consequently, the results of the study may reflect the rating tendencies of these specific teachers as much as the work habits of the students rated. Teachers not only rated multiple students, some teachers rated multiple classes of students. Using teachers as raters of student characteristics typically necessitates that one teacher will rate multiple students. However, to increase the diversity of raters, perhaps one rater should be restricted to one class of students. Even better would be to have multiple teachers rate the same students in different classes, permitting cross-validation of teacher ratings. A possible problem with the latter approach, however, is that work habits might be somewhat situation specific, causing teachers in different classes to rate the student's work habits differently.

Although internal consistency is generally viewed as more fundamental to the psychometric adequacy of a scale than is test-retest consistency, the latter metric is important in the further development of the Student Work Habits Rating Form. Because both student and teacher behaviors and perspectives may vary considerably from day to day, the stability of the work-habits ratings needs to be established. Would the same

student be rated differently on a good day or a bad day? How the student behaves on a particular day could influence teacher ratings of more enduring tendencies. Perhaps as important as daily variation in student behavior is the teacher's day-to-day variation in morale and effectiveness. Teacher ratings on a bad day for him or her could be quite different from that teacher's ratings on a good day. Then again, teachers may be able to transcend their own mood variations in assessing a student's typical work habits. Only assessment of test-retest consistency in teacher ratings of the same students will answer such questions.

The selection of criterion measures are exceedingly important in the psychometric development of such instruments as the Student Work Habits Rating Form. The criterion measures in the current study were extremely limited. In fact, only a 3-category teacher rating of student work habits was used as a criterion measure. More diverse measures need to be included (e.g., overall GPA, standardized test scores, performance on projects, observational measures of on-task behavior, and student self-rating of work habits); more precise measures need to be used (e.g., a full range of grades as compared to a 3-category rating); and the criterion measure needs to be largely independent of the teachers' ratings. If the same teachers rate both the work habits and the criterion measure, there is ample opportunity for one set of ratings to influence the other set of ratings. Rather than the arrangement found in the current study, it would be preferable to use criterion measures from sources other than the teachers who do the ratings of work habits (e.g., standardized test scores, overall GPA).

Conclusions

The Student Work Habits Rating Form is a work in progress. The results thus far are sufficiently promising to warrant further development of the rating form. Currently, the instrument has high internal consistency and moderate criterion-related validity. It is best considered to be a general measure of student work habits. Much work remains to be done before the instrument can be recommended as a practical tool in making judgments about individual students and interventions for promoting good work habits. Additional work habits need to be added to the form, and a better balance needs to be achieved between positive and negative work habits. A greater diversity of teacher raters needs to be included in the followup research on the instrument, with some provision for cross-validation of teacher ratings. Finally, the student sample needs to be expanded to include more representative sampling of different ethnic groups. With continued research along the lines suggested in this chapter, the Student Work Habits Rating Form could become a psychometrically sound measure of a construct (work habits) that might be fundamental to the long-term success of middle school students.

When the recommended research has been done on the Student Work Habits Rating Form, it could be useful for counselors and teachers who want to better understand their students' work habits. When using this scale, educators may realize that the problems of poor academic students relate not only to the content of the material but also to students' work habits. This scale could help educators identify the specific work habits that are fundamental to students' academic achievement.

REFERENCES

REFERENCES

- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443–459.
- Book, W. F. (1926). *Learning how to study and work effectively*. Boston: Ginn and Company.
- Boveja, M. E. (1998). Parenting styles and adolescents' learning strategies in the urban community. *Journal of Multicultural Counseling and Development*, 26, 110–119.
- Bush, V. (1999). *Development of a work-ethic scale for middle school students*. Unpublished doctoral dissertation, The University of Tennessee, Knoxville.
- Cobb, J. A. (1972). The relationship of discrete classroom behaviors to fourth-grade academic achievement. *Journal of Educational Psychology*, 63, 74–80.
- Cooper, H. M., & Lowe, C. A. (1977). Task information and attributions for academic performance by professional teachers and role players. *Journal of Personality*, 45, 469–483.
- Cottier, S. J., & Koehler, S. A. (1978). A study skills unit for junior high students. *Journal of Reading*, 21, 626–630.
- Davenport, E., Spiegel, E. J., Klein, P., Fitzpatrick, E. M., & Mueller, P. (1984). Study skills: Tools of the trade to make studying easier and more efficient. *Early Years*, 15, 43–52, 81.

- DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.
- Devine, T. G. (1987). *Teaching study skills: A guide for teachers* (2nd ed.). Boston: Allyn and Bacon.
- Diener, C. L. (1960). Similarities and differences between over-achieving and under-achieving students. *Personnel and Guidance Journal*, 38, 396–400.
- Doherty, J., & Conolly, M. (1985). How accurately can primary school teachers predict the scores of their pupils in standardized tests of attainment? A study of some non-cognitive factors that influence specific judgements. *Educational Studies*, 11, 41–60.
- Dudycha, G. J. (1936). An objective study of personality. *Archives of Psychology Columbia University*, 204, 53.
- Edelbrock, C. (1983). Problems and issues in using rating scales to assess child personality and psychopathology. *School Psychology Review*, 12, 293–299.
- Ediger, M. (1987). *School dropouts, absenteeism, and tardiness*. (EDRS No. RIEAUG87).
- Ehman, L. H. (1970). *A comparison of three sources of classroom data: Teachers, students, and systematic observation*. Paper presented at the annual meeting of the American Educational Research Association, Minneapolis. (ERIC Document Reproduction Service No. ED 037 404)
- Elliott, D. S., Voss, H. L., & Wendling, A. (1966). Capable dropouts and the social milieu of the high school. *Journal of Educational Research*, 60, 180–186.

- Erdman, A. (1992, August 10). What's wrong with workers. *Fortune*, p.18.
- Evans, E. D., & Tribble, M. S. (1987). Study skills measurement in grades 7-12. *Journal of Adolescent Research*, 2(1), 85-93.
- Fagot, B. I. (1973). Influence of teacher behavior in the preschool. *Developmental Psychology*, 9, 198-206.
- Farkas, G., Grobe, R. P., Sheehan, D., & Shuan, Y. (1990). Cultural resources and school success: Gender, ethnicity, and poverty groups within an urban school district. *American Sociological Review*, 55, 127-142.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Forness, S. R., Guthrie, D., & Hall, R. J. (1976). Follow-up of high-risk children identified in kindergarten through direct classroom observation. *Psychology in the Schools*, 13(1), 45-49.
- Fuchs, L. S., Fuchs, D., & Phillips, N. (1994). The relation between teachers' beliefs about the importance of good student work habits, teacher planning, and student achievement. *The Elementary School Journal*, 94, 331-345.
- Gadzella, B. M., & Williamson, J. D. (1984). Study skills, self-concept, and academic achievement. *Psychological Reports*, 54, 923-929.
- Georgiady, N. P., & Romano, L. C. (1994). *Focus on study habits in school: A guide for teachers and students to increase learning in the middle school* (Michigan Association of Middle School Educators, ISBN-0-918449-06-5). East Lansing: Michigan State University.

- Goodstein, M. (1979). *The diagnosis and treatment of underachievement*. Paper presented at the B'nai B'rith Career and Counseling Services Professional Staff Conference. Washington, DC: B'nai B'rith. (ERIC Document Reproduction Service No. ED 184 036)
- Graham, K. G., & Robinson, H. A. (1984). *Study skills handbook: A guide for all teachers*. Newark, DE: International Reading Association.
- Greene, L. J. (1986). *Kids who underachieve*. New York: Simon and Schuster.
- Gronlund, N. E. (1976). *Measurement and evaluation in teaching* (3rd ed.). New York: Macmillan.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- Haberman, M. (1997, March). Unemployment training: The ideology of nonwork learned in urban schools. *Phi Delta Kappan*, 78, 499–503.
- Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21, 967–988.
- Hirsch, V. (1983). *The validity and reliability of the Brown-Gadzella CAI study skills test*. Unpublished doctoral dissertation, East Texas State University, Texarkana.

- Hurlburt, G., Gade, E., & Fuqua, D. (1985). *Study habits and attitudes of Indian students: Implications for counselor involvement*. (ERIC Document Reproduction Service No. ED 272 349).
- Jecker, J., Maccoby, N., Breitrose, H. S., & Rose, E. D. (1964). Teacher accuracy in assessing cognitive visual feedback from students. *Journal of Applied Psychology*, 48, 393–397.
- Jöreskog, K. G., & Sörbom, D. (1993a). *PRELIS 2 user's reference guide*. Chicago: Scientific Software International.
- Jöreskog, K. G., & Sörbom, D. (1993b). *LISREL 8 user's guide*. Chicago: Scientific Software International.
- Kamphaus, R. W., & Frick, P. J. (1996). *Clinical assessment of child and adolescent personality and behavior*. Boston, Allyn and Bacon.
- Kazanas, H. C., & Wolff, L. C. (1972). Development of work habits in vocational education—What the literature indicates. *Journal of Industrial Teacher Education*, 10, 48–58.
- Keith, T. Z., Reimers, T. M., Fehrmann, P. G., Pottebaum, S. M., & Aubey, L. W. (1986). Parental involvement, homework, and TV time: Direct and indirect effects on high school achievement. *Journal of Educational Psychology*, 78, 373–380.
- Kerman, S. (1979). Teacher expectations and student achievement. *Phi Delta Kappan*, 60, 716–718.

- Kirisci, L., Moss, H. B., & Tarter, R. E. (1996). Psychometric evaluation of the situational confidence questionnaire in adolescents: Fitting a graded item response model. *Addictive Behaviors, 21*, 303–317.
- Kitson, H. D. (1916). *How to use your mind*. Philadelphia: J. B. Lippincott.
- Kline, P. (1986). *A handbook of test construction: Introduction to psychometric design*. New York: Methuen.
- Koch, W. R. (1983). Likert scaling using the graded response latent trait model. *Applied Psychological Measurement, 7*(1), 15–32.
- Kops, C., & Belmont, I. (1985). Planning and organizing skills of poor achievers. *Journal of Learning Disabilities, 18*(1), 8–14.
- Levine, F. J. (1976). Influence of field-independence and study habits on academic performance of Black students in a predominantly White university. *Perceptual and Motor Skills, 42*, 1101–1102.
- Locke, E. A. (1977). An empirical study of lecture note taking among college students. *Journal of Educational Research, 71*, 93–99.
- Lynch, A. M. (1997). *The application of item response theory to employee attitude survey data using Samejima's Graded Response Model*. Unpublished doctoral dissertation, The University of Connecticut, Storrs.
- Malone, Y. (1998). *Development of an instrument to measure middle schoolers' perceptions of the relationship between school-related behaviors and adult opportunities*. Unpublished doctoral dissertation, The University of Tennessee, Knoxville.

- Mandel, H. P., & Marcus, S. I. (1995). *"Could Do Better": Why children underachieve and what to do about it*. New York: John Wiley & Sons.
- Messick, S. (1962). Response style and content measures from personality inventories. *Educational and Psychological Measurement*, 22, 41–56.
- Meyers, C. E., Attwell, A. A., & Orpet, R. E. (1968). Prediction of fifth grade achievement from kindergarten test and rating data. *Educational and Psychological Measurement*, 28, 457–463.
- Natriello, G., & Dornbusch, S. M. (1983). Bringing behavior back in: The effects of student characteristics and behavior on the classroom behavior of teacher. *American Educational Research Journal*, 20(1), 29–43.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- O'Leary, M. J. (1955). *The measurement and evaluation of the work habits of over-achievers and under-achievers to determine the relationship of these habits to achievement*. Unpublished doctoral dissertation, Boston University, Boston.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.
- Prater, M. A., Hogan, S., & Miller, S. R. (1992). Using self-monitoring to improve on-task behavior and academic skills of an adolescent with mild handicaps across special and regular education settings. *Education and Treatment of Children*, 15, 43–55.

- Rafoth, M. A., & DeFabo, L. (1990). *Study skills: What research says to the teacher*. National Education Association, Washington, D.C. (ERIC Document No. ED 323 184)
- Rao, D. G. (1973). A scale to measure the study habits of high school students. *MANAS*, 20, 143–149.
- Raph, J. B., Goldberg, M. L., & Passow, A. H. (1966). *Bright underachievers*. New York: Teachers College Press, Columbia University.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.
- Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 85–100). New York: Springer-Verlag.
- Schreiber, D. (1962). School dropouts. *National Education Association Journal*, 51, 50–59.
- Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, 41, 1101–1114.
- Schroeder, T. S. (1978). *Children's work habits and reading performance*. Paper presented at the Annual Meeting of the International Reading Association. (ERIC Document Reproduction Service No. ED 155 669)

- Scott, J. L. (1973). *The effects of a learning program on work habit skills and attitudes toward work*. Unpublished doctoral dissertation, The University of Missouri, Columbia.
- Sears, S. J. (1985, April). Career and educational planning in middle level school. *NASSP Bulletin*, 37–42.
- Soli, S. D., & Devine, V. T. (1976). Behavioral correlates of achievement: A look at high and low achievers. *Journal of Educational Psychology*, 68, 335–341.
- Spaulding, C. L. (1992). *Motivation in the classroom*. New York: McGraw-Hill.
- Stipek, D. J. (1983). Work habits begin in preschool. *Young Children*, 38(4), 25–32.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins College Publishers.
- Takei, Y., Johnson, M. P., & Clark, M. E. (1998). Academic achievement and impression management as factors in the grading of white junior high pupils. *Sociological Perspectives*, 41(1), 27–48.
- Thissen, D. (1991). *Multilog user's guide*. Chicago: Scientific Software.
- Tollefson, N., Melvin, J., & Thippavajjala, C. (1990). Teachers' attributions for students' low achievement: A validation of Cooper and Good's attributional categories. *Psychology in the Schools*, 27, 75–83.
- Traxler, A. E. (1954). The improvement of study habits and skills. *Educational Records Bulletin (Revised)*, 41. New York: Educational Records Bureau.
- Turney, A. H. (1930). *Factors other than intelligence that affect success in high school*. Minneapolis: University of Minnesota Press.

- Uttaro, T., & Lehman, A. (1999). Graded response modeling of the quality of life interview. *Evaluation and Program Planning*, 22(1), 41–52.
- Weigel, R. G., & Weigel, V. M. (1967). The relationship of knowledge and usage of study skill techniques to academic performance. *Journal of Educational Research*, 61(2), 78–80.
- Weinland, J. D. (1930). How successful college students study. *Journal of Educational Psychology*, 31, 521–526.
- Weitzman, M. (1985). Demographic and educational characteristics of inner-city middle school problem absence students. *American Journal of Orthopsychiatry*, 55, 378–383.
- Wells, R. H. (1991). *Build on-the-job success skills. Breakthrough strategies to teach and counsel troubled youth: Social skills, schools skills, coping skills lesson series*. Woodburn, OR: Youth Change, RIEAPR94.
- Wheldall, K., & Panagopoulou-Stamatelatou, A. (1991). The effects of pupil self-recording of on-task behaviour on primary school children. *British Educational Research Journal*, 17, 113–127.
- Zarb, J. M. (1981). Non-academic predictors of successful academic achievement in a normal adolescent sample. *Adolescence*, 16, 891–900.

APPENDICES

APPENDIX A

ACHIEVEMENT RATING FORM

School Number _____ Grade Level _____ Class Number _____

INSTRUCTIONS: Using the corresponding number in your roll book to identify each student, please rate each student on the basis of that student's past academic performance in your class. Check the column below that best describes the academic performance of each student. Rate all the students in each class, even those who do not take the Middle School Survey. However, for those who do take the Middle School Survey, be sure that their number on this rating form is consistent with the number on the student answer sheet.

Student Number	Poor Achievement	Average Achievement	Outstanding Achievement
1.	_____	_____	_____
2.	_____	_____	_____
3.	_____	_____	_____
4.	_____	_____	_____
5.	_____	_____	_____
6.	_____	_____	_____
7.	_____	_____	_____
8.	_____	_____	_____
9.	_____	_____	_____
9.	_____	_____	_____
10.	_____	_____	_____
11.	_____	_____	_____
12.	_____	_____	_____
13.	_____	_____	_____
14.	_____	_____	_____
15.	_____	_____	_____
16.	_____	_____	_____
17.	_____	_____	_____

Student Number	Poor Achievement	Average Achievement	Outstanding Achievement
18.	_____	_____	_____
19.	_____	_____	_____
20.	_____	_____	_____
21.	_____	_____	_____
22.	_____	_____	_____
23.	_____	_____	_____
24.	_____	_____	_____
25.	_____	_____	_____
26.	_____	_____	_____
27.	_____	_____	_____
28.	_____	_____	_____
29.	_____	_____	_____
30.	_____	_____	_____
31.	_____	_____	_____
32.	_____	_____	_____
33.	_____	_____	_____
34.	_____	_____	_____
35.	_____	_____	_____
36.	_____	_____	_____
37.	_____	_____	_____
38.	_____	_____	_____
39.	_____	_____	_____
40.	_____	_____	_____
41.	_____	_____	_____
42.	_____	_____	_____
43.	_____	_____	_____
44.	_____	_____	_____
45.	_____	_____	_____

APPENDIX B

STUDENT WORK HABITS RATING FORM

School Number____ Grade Level____ Class Number____ Student Number____

Check One: Best Work Habits____ or Worst Work Habits____

INSTRUCTIONS: For the student identified above, please rate that student on each of the characteristics listed below. To help us understand how student work habits may be linked to attitudes and performance, you should be very candid in rating the work habits of this student. To do this, think of a typical day for the student and indicate whether the work habit in question is (1) a lot like this student, (2) somewhat like this student, (3) neither like nor unlike this student, (4) somewhat unlike this student, or (5) very unlike this student. Please put your rating number in the blank prior to each statement.

- ____ 1. Follows rules regarding classroom behavior.
- ____ 2. Cooperates with other students.
- ____ 3. Needs much supervision.
- ____ 4. Arrives at class on time.
- ____ 5. Easily gets upset.
- ____ 6. Adapts easily to different kinds of school work.
- ____ 7. Keeps promises.
- ____ 8. Expresses needs clearly and comfortably.
- ____ 9. Takes good care of books and other classroom materials.
- ____ 10. Addresses teachers in a respectful fashion.
- ____ 11. Practices good attendance.
- ____ 12. Keeps work space well organized.
- ____ 13. Expresses enthusiasm about school work.
- ____ 14. Takes initiative to keep busy.
- ____ 15. Pestors other students.
- ____ 16. Makes good use of teacher feedback.
- ____ 17. Does good quality work.
- ____ 18. Follows instructions well.
- ____ 19. Works at a steady pace.
- ____ 20. Is easily distracted.

APPENDIX C

Inter-Item Correlation Matrix Polychoric Correlation (Below Diagonal) and Pearson Product Moment Correlation (Above Diagonal)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1		0.813	0.641	0.698	0.446	0.670	0.724	0.645	0.748	0.753	0.435	0.661	0.681	0.717	0.579	0.739	0.724	0.748	0.711	0.610
2	0.895		0.643	0.656	0.539	0.733	0.761	0.676	0.715	0.709	0.397	0.649	0.693	0.753	0.606	0.732	0.724	0.777	0.726	0.610
3	0.799	0.778		0.505	0.466	0.641	0.626	0.534	0.658	0.480	0.309	0.614	0.637	0.681	0.603	0.638	0.656	0.714	0.686	0.648
4	0.822	0.779	0.678		0.374	0.602	0.600	0.543	0.639	0.605	0.474	0.501	0.533	0.592	0.521	0.562	0.590	0.617	0.591	0.481
5	0.581	0.657	0.603	0.511		0.515	0.470	0.421	0.462	0.527	0.267	0.414	0.435	0.526	0.449	0.481	0.461	0.516	0.495	0.546
6	0.785	0.839	0.774	0.758	0.649		0.741	0.766	0.771	0.591	0.518	0.770	0.807	0.827	0.564	0.819	0.849	0.827	0.842	0.703
7	0.840	0.856	0.764	0.759	0.602	0.845		0.726	0.748	0.699	0.479	0.745	0.732	0.773	0.612	0.780	0.786	0.794	0.770	0.642
8	0.748	0.773	0.663	0.686	0.542	0.840	0.821		0.745	0.605	0.573	0.750	0.764	0.786	0.482	0.793	0.807	0.763	0.758	0.609
9	0.843	0.818	0.782	0.780	0.586	0.850	0.847	0.831		0.691	0.515	0.858	0.787	0.831	0.593	0.798	0.835	0.841	0.844	0.678
10	0.858	0.819	0.670	0.738	0.682	0.744	0.832	0.733	0.813		0.407	0.615	0.550	0.668	0.543	0.657	0.644	0.659	0.635	0.554
11	0.589	0.558	0.483	0.649	0.428	0.690	0.644	0.729	0.691	0.583		0.502	0.494	0.503	0.323	0.472	0.493	0.476	0.471	0.377
12	0.766	0.757	0.730	0.660	0.513	0.840	0.838	0.827	0.903	0.733	0.662		0.786	0.825	0.544	0.810	0.857	0.842	0.842	0.682
13	0.808	0.799	0.756	0.704	0.539	0.874	0.831	0.862	0.877	0.715	0.668	0.860		0.869	0.540	0.854	0.894	0.873	0.863	0.707
14	0.849	0.871	0.817	0.789	0.663	0.900	0.874	0.873	0.918	0.843	0.704	0.898	0.936		0.576	0.874	0.902	0.877	0.914	0.745
15	0.726	0.744	0.716	0.677	0.552	0.699	0.734	0.576	0.694	0.698	0.501	0.659	0.634	0.701		0.544	0.578	0.628	0.618	0.645
16	0.840	0.830	0.761	0.722	0.608	0.884	0.863	0.875	0.869	0.795	0.632	0.871	0.918	0.932	0.659		0.906	0.876	0.891	0.690
17	0.845	0.848	0.775	0.759	0.593	0.911	0.878	0.887	0.903	0.801	0.661	0.918	0.935	0.951	0.712	0.951		0.919	0.932	0.753
18	0.863	0.875	0.823	0.778	0.649	0.903	0.880	0.849	0.908	0.814	0.645	0.893	0.924	0.940	0.756	0.935	0.963		0.913	0.743
19	0.820	0.842	0.811	0.760	0.619	0.904	0.865	0.841	0.919	0.791	0.629	0.905	0.911	0.958	0.739	0.936	0.968	0.963		0.765
20	0.789	0.776	0.792	0.661	0.682	0.830	0.780	0.740	0.804	0.745	0.544	0.804	0.800	0.852	0.769	0.802	0.863	0.866	0.873	

APPENDIX D

Parameter Estimates and Item- and Test-Information Functions

Item	Discrimination	Difficulty*				θ : Latent Trait Level								
	<i>a</i> (s.e.)	<i>b</i> 1	<i>b</i> 2	<i>b</i> 3	<i>b</i> 4	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0
1	2.67 (.35)	-1.20	-.62	-.50	0.29	0.67	1.56	2.12	2.16	2.03	1.70	0.81	0.26	0.07
2	2.74 (.33)	-1.26	-.63	-.40	0.42	0.78	1.75	2.23	2.30	2.13	1.94	1.05	0.35	0.10
3	2.00 (.22)	-.64	-.09	0.11	0.54	0.23	0.52	0.92	1.20	1.27	1.17	0.84	0.45	0.19
4	1.79 (.29)	-1.98	-1.25	-.92	-.21	0.90	1.00	1.01	0.97	0.83	0.56	0.30	0.14	0.06
5	1.24 (.19)	-1.76	-.73	-.07	0.58	0.41	0.46	0.49	0.49	0.49	0.45	0.39	0.29	0.19
6	3.09 (.35)	-.80	-.29	-.05	0.53	0.22	0.88	2.26	2.93	2.96	2.68	1.48	0.43	0.10
7	3.24 (.41)	-.99	-.59	-.25	0.51	0.37	1.42	2.95	3.30	2.94	2.82	1.49	0.39	0.08
8	2.42 (.26)	-.98	-.49	-.20	0.56	0.42	1.02	1.65	1.85	1.78	1.65	1.14	0.50	0.17
9	3.39 (.42)	-.87	-.46	-.20	0.49	0.24	1.08	2.90	3.60	3.32	3.07	1.47	0.35	0.07
10	2.26 (.32)	-1.65	-1.06	-.65	-.07	1.13	1.52	1.62	1.60	1.40	0.87	0.38	0.14	0.05
11	1.23 (.23)	-2.50	-1.80	-1.46	-.43	0.47	0.48	0.46	0.43	0.37	0.28	0.19	0.12	0.07
12	2.90 (.34)	-.96	-.31	-.15	0.44	0.37	1.20	2.26	2.53	2.57	2.26	1.16	0.35	0.09
13	2.98 (.34)	-.32	0.06	0.28	0.94	0.06	0.25	0.92	2.19	2.78	2.64	2.35	1.20	0.35
14	4.11 (.43)	-.34	-.07	0.13	0.85	0.02	0.14	0.98	3.96	5.25	4.17	3.91	1.02	0.15
15	1.58 (.19)	-.93	-.26	0.08	0.57	0.33	0.53	0.70	0.78	0.79	0.73	0.58	0.38	0.21
16	3.52 (.41)	-.60	-.21	0.03	0.67	0.09	0.47	1.97	3.69	3.82	3.48	2.25	0.59	0.11
17	5.14 (.71)	-.45	-.03	0.11	0.54	0.01	0.12	1.37	6.81	8.06	7.29	2.06	0.19	0.01
18	5.51 (.83)	-.58	-.24	0.02	0.43	0.01	0.19	2.49	8.76	9.12	7.61	1.23	0.08	0.01
19	4.39 (.55)	-.48	-.05	0.11	0.52	0.02	0.21	1.63	5.23	5.98	5.40	1.86	0.25	0.03
20	2.45 (.26)	-.28	0.28	0.53	0.87	0.09	0.27	0.75	1.47	1.85	1.90	1.60	0.88	0.33
Test-Information Function														
						6.84	15.07	31.68	56.25	59.74	52.67	26.54	8.36	2.44

**b* values represent threshold of responses between Likert options. For example, *b* 1 is the threshold of responses between option 1 and option 2.

VITA

Eun Jung Oh was born in Seoul, South Korea. She attended public schools in Sungnam, South Korea. She received a Bachelor of Arts in History Education from Chong Ju University, South Korea, in 1992. She entered the Master's program in Guidance and Counseling at The University of Tennessee, Chattanooga in August of 1994, receiving the Master of Education degree in 1996. In August of 1996, she entered The University of Tennessee, Knoxville, to pursue the Doctor of Philosophy degree in Educational Psychology.