

Reliability of the Murphy-Meisgeier Type Indicator for Children (MMTIC) and the
Student Styles Questionnaire (SSQ)

Isaac Fitzsimons

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Education

University of Washington

2014

Committee:

Deborah McCutchen

Karin Frey

Dixie Massey

Program Authorized to Offer Degree:

Education

© Copyright 2014
Isaac Fitzsimons

University of Washington

Abstract

Reliability of the Murphy-Meisgeier Type Indicator for Children (MMTIC) and the
Student Styles Questionnaire (SSQ)

Isaac Fitzsimons

Chair of the Supervisory Committee:
Dr. Deborah McCutchen
College of Education

Research suggests that psychological type affects one's status in social groups. This, in turn, may impact performance in the classroom when students are required to participate in cooperative learning. At the same time, research on the psychometric properties of instruments used to measure psychological type in children and adolescents is almost non-existent. In order to decrease this gap between theory and practice, a systematic literature review was conducted for studies that used the Murphy-Meisgeier Type Indicator for Children (MMTIC) or the Student Styles Questionnaire (SSQ) to measure psychological type in children and adolescents. Reliability estimates were extracted and compared in order to determine which instrument provided adequate reliability. Results clearly indicated that the revised MMTIC (2008) yielded the strongest internal consistency and test-retest reliability estimates, although differences in reliability existed between scales.

Introduction

Creating safe classrooms where all students are given the same opportunities to participate in classroom activities is a priority for most teachers. Indeed, many studies have shown that students who tend to participate have more positive learning outcomes than students who do not (Bianchini, 1997; Cohen, 1994; Slavin, Hurley, & Chamberlain, 2003; Webb, 1989). Experience has demonstrated however, that even with safe learning ecosystems, status hierarchies can and do occur where certain students have a greater influence, contribute more, and are considered more competent than others (Cohen & Lotan, 1995). There is evidence that the presence of status hierarchies in the classroom can negatively affect both student learning and student social-emotional health (Cohen, 1994; Garandau, Lee, & Salmivalli, 2013). Cooperative learning is one of many instructional strategies used by teachers to decrease the effects of status hierarchies in their classrooms. However, even in classrooms where teachers use cooperative learning techniques designed to reduce status, status hierarchies can still occur and the effectiveness of cooperative learning is reduced (Bianchini, 1997).

Researchers hypothesize that status hierarchies form when individual attributes, called *status characteristics*, are used to determine the competence of group members, even when these attributes do not reflect actual ability to complete the task (Melamed & Savage, 2013). Race and gender are examples of two status characteristics that have been shown to affect student interactions (Cohen, 1994; Katz & Benjamin, 1960; Webb, 1984). Research in the workplace has used the Myers-Briggs Type Indicator (MBTI) to study personality as another status characteristic that influences interactions in groups (Ostafichuk & Naylor, 2013). However there is little research on personality as a status characteristic that affects student performance during cooperative learning in the classroom.

The MBTI is a popular instrument used to measure psychological type in adult non-psychiatric populations (Furnham, 1996) and research has been done to confirm its reliability (e.g., Capraro & Capraro, 2002). Meanwhile, there is less research concerning the psychometric properties of instruments used to measure psychological type in children. Confidence in the psychometric properties of these instruments is necessary for researchers who use them in an educational setting with children and adolescents.

In the following review, I will examine the reliability of two instruments that are commonly used to measure psychological type in children and adolescents in order to contribute to research on their psychometric properties. This will benefit the field of education, as educational researchers will be able to choose the appropriate instrument to assess psychological type. Both instruments will be discussed at length later in this paper. Since psychological type affects status in social groups, I will first review relevant literature on the effects of status on student interactions in the classroom. Then, I will consider four perspectives on cooperative learning and how status affects student outcomes. Next, I will expand on why psychological type could be considered a status characteristic and continue with a discussion about psychological type and how it is measured in children and adolescents. I will then touch on current issues with the psychometric properties of the instruments used. Finally, I will end by explaining the importance of establishing reliability in instruments. By contributing to research on the psychometric properties of instruments used to measure psychological type in children and adolescents, this paper will help researchers better understand personality as a status characteristic that influences student access to participation in cooperative learning, which will help teachers to create more equitable classrooms.

Literature Review

Status

Status refers to one's position in relation to the rest of society (Linton, 1936). Being ranked high in status is considered better than being ranked low in status (Cohen, 1994). Status is determined using status characteristics, which are defined as "... attribute(s) of an individual with widely held cultural beliefs that attach greater value and competence to one category of the attribute than the other" (Melamed & Savage, 2013, p. 1086). Status characteristics can be further separated into two categories: *ascribed* and *acquired* (Linton, 1936). Ascribed status characteristics are those that are inborn and immutable, such as gender or race (Linton, 1936). Acquired status characteristics, also known as achieved status characteristics, are those that one gains through experience and effort, such as education (Linton, 1936).

Ascribed and acquired status characteristics influence the interactions of students in both laboratory and classroom settings (Cohen, 1994; Katz & Benjamin, 1960; Webb, 1984). In the classroom, academic competence is considered to be the most salient acquired status characteristic. It is well documented that students labeled low in academic competence interact less frequently and have less influence in small group discussions than students who are labeled high in academic competence (Cohen, 1994; Dembo & McAuliffe, 1987; Rosenholtz, 1985). This effect has been found even when academic competence was determined using a bogus achievement test, suggesting that perceived academic competence is not necessarily associated with actual academic competence (Dembo & McAuliffe, 1987).

Ascribed status characteristics, such as race and gender, also influence student interactions. For example, when working on a cooperative task, African American college students were socially inhibited and submissive towards their white partners, even when they

were matched on intelligence and ability (Katz & Benjamin, 1960). Additionally, when involved in a group task, white junior high boys were more active and influential than black junior high boys (Cohen, 1994). Gender also affects group interactions. For instance, when working on math problems in cooperative groups that consisted mostly of males, the contributions of female students were ignored (Webb, 1984). Furthermore, even when the groups consisted mostly of females, the female students tended to defer to the male students, who then did most of the participation (Webb, 1984). Ascribed and acquired status characteristics also interact with one another as classroom popularity is often highly correlated with perceived academic status, and academic status is often determined using ascribed status characteristics (Rosenholtz & Wilson, 1980).

Formation of Status Hierarchies. Status hierarchies form when status characteristics are used to create a ranking based on perceived competence (Melamed & Savage, 2013). Status hierarchies are thought to have some evolutionary benefit as they enhance the predictability and stability of social relationships, which can create social order and improve harmony (Garandau et al., 2013). However, the presence of status hierarchies in classrooms has been linked to both social-emotional and academic problems (Cohen, 1994; Garandau et al., 2013; Wilson, Karimpour, & Rodkin, 2011). Indeed, student status is correlated with the amount of interaction and participation in classroom discussions, which in turn is a predictor of learning gains (Cohen, 1994). Additionally, 8th grade classrooms with a higher degree of status hierarchy resulted in greater rates of bullying at the end of the year compared to classrooms with a lower degree of status hierarchy (Garandau et al., 2013). Further, in a longitudinal study, girls with low social status in 6th grade had higher rates of hospitalization for anxiety and depression as adults than girls with average social status (Modin, Ostberg, & Almquist, 2011). Unexpectedly, girls who

were high in status in 6th grade were also at an increased, but not statistically significant, risk of being hospitalized for anxiety and depression when compared to girls with average social status (Modin et al., 2011). It seems clear, therefore, that being in classrooms where students have equal status is beneficial for learning outcomes and social-emotional health.

In classrooms, status hierarchies are created using several dimensions such as popularity, attractiveness, and academic competence in the particular subject (Jonkmann, Trautwein, & Ludtke, 2009). Academic competence is often evaluated by *communicative status*, which is gained through participation (Morine-Dershimer, 1983). Teacher expectation for student discourse in the classroom can affect which students participate, and this, in turn, has implications for the formation of status hierarchies (Morine-Dershimer, 1983). For example, in a classroom where the teacher expected full student participation and asked questions with simple answers (easily found in the textbook), all students, regardless of actual academic competence, contributed (Morine-Dershimer, 1983). In other words, all students gained equal communicative competence. Conversely, in a classroom where the expectation was to contribute high quality responses, only students high in actual academic competence could contribute, creating a status hierarchy with students high in academic competence at the top and students low in academic competence at the bottom (Morine-Dershimer, 1983). Finally, in a classroom where the expectation was for students to contribute divergent ideas, students low in status, as measured by actual academic competence and social status, contributed more, ostensibly because they felt that they had less to lose than high status students if they made a mistake (Morine-Dershimer, 1983). All three classrooms displayed significant weaknesses in terms of student participation. In the first classroom, there was equality of participation, but the comments made were not valuable. In the second classroom, there was inequality of participation, which contributed to the formation

of the status hierarchy. In the third classroom, there was also inequality of participation, but because the students low in academic competence gained communicative status, the classroom did not benefit from the knowledge of students high in academic competence.

Teachers can influence the formation of status hierarchies in their classrooms by using specific instructional strategies (Cohen, 1994; Morine-Dersheimer, 1983). In a multiple abilities treatment teachers convince students that many different abilities are required to complete a group task (Cohen, 1994; Rosenholtz, 1985). Additionally, teachers can raise the status of low status students by assigning competence, where they publicly praise students low in status when they display competence on an important intellectual ability (Cohen, 1994). These instructional strategies are generally used in conjunction with *cooperative learning*, an instructional approach in which individuals work together in small groups to solve problems and help each other learn (Webb, Farivar, & Mastergeorge, 2001). In this way, educators have used cooperative learning to create more equitable student interactions in the classroom (Cohen, 1994).

Cooperative Learning

Research on cooperative learning originated not in education but in psychology (Hogan & Tudge, 1999), but was widely used as an instructional strategy in the “progressive” era of education and was considered a valuable educational tool by John Dewey and other educational theorists. Cooperative learning fell out of favor in the 1960s when schools began emphasizing competition and individualism (Johnson & Johnson, 2009). In the 1980s, it began to regain popularity again and is continuing to be used today (Johnson & Johnson, 2009). The following four theoretical perspectives have contributed to research on cooperative learning: the *social interdependence perspective*, the *motivationalist perspective*, the *cognitive developmental perspective*, and the *cognitive elaboration perspective*. The development and essential details of

each perspective are described below.

Social Interdependence Perspective. The social interdependence theory, also known as the social cohesion theory, is considered to be one of the most successful and widespread applications of social and educational psychology into practice (Johnson & Johnson, 2009). According to the theory, interdependence occurs when people believe that their outcomes are affected by others' actions as well as their own. Interdependence can be positive, meaning that people believe they can only achieve their individual goal if others in the group are also able to achieve their own goals. This application results in helping others to achieve their goals (Johnson & Johnson, 2009). On the other hand, interdependence can also be negative, meaning that people believe that they can only achieve their individual goals if others are prevented from achieving their own goals. This application results in the obstruction of others' goals (Johnson & Johnson, 2009).

According to social interdependence theory, cooperative learning is a successful instructional strategy when students are motivated to help one another learn. Theorists believe that the effectiveness of a cooperative learning exercise depends on the cohesiveness of the group, or how much students identify as members of the group (Slavin et al., 2003). Social interdependence theorists emphasize that teachers must prepare students for group work by using team-building activities in order to create interdependence among members. Some support exists to suggest that achievement in cooperative learning depends on the quality of group interactions and that group interactions depend on the social cohesion of the group. However, there is inconsistent evidence to suggest that team-building exercises alone enhance student achievement in cooperative learning (Slavin et al., 2003).

Motivationalist Perspective. The second perspective, the motivationalist perspective,

emphasizes the role that task motivation plays in the learning process. Motivational theorists believe that in order for cooperative learning to be successful, groups must have rewards or goals. The motivationalist perspective is similar to social interdependence theory in that both are concerned with the motivation of individual group members to complete a group task. However, this perspective states that goals and rewards, not social cohesion, are what motivate students. According to this perspective, traditional classrooms that use competitive grading and rankings undermine students' motivation to help one another learn. Similar to the social interdependence theory, there is mixed support for this perspective. For example, group goals and group rewards only result in increased academic achievement if group members are also required to demonstrate individual learning; otherwise one or two group members may do all the work (Slavin et al., 2003).

Cognitive Perspectives. The cognitive developmental and cognitive elaboration perspectives fall under a larger cognitive umbrella. Cognitive theorists, such as Vygotsky and Piaget, believe that it is interaction among group members that leads to better mental processing, and that motivation has little to no effect on group performance (Slavin et al., 2003). Vygotsky further claimed that when children engage in problem solving with someone who has greater competence, both participants gain knowledge (Hogan & Tudge, 1999). There are two well-known components of Vygotsky's theory: the *zone of proximal development* (ZPD) and *intersubjectivity*. The zone of proximal development is defined as the difference between what a person can accomplish independently and what they are able to accomplish with the help of a more competent partner (Vygotsky, 1978). Another way to conceptualize the ZPD is as children's developmental potential that can be exposed through working with others. Once children uncover their potential, they can internalize those processes and use them independently

(Vygotsky, 1978). For Vygotsky, intersubjectivity occurs when participants approach a problem-solving task from differing viewpoints but through collaboration, arrive at a shared understanding.

Other educators also espoused benefits of working with others. Piaget argued that certain types of knowledge, such as language, values, and rules, could only be learned through interactions with others (Slavin et al., 2003). Even knowledge that children may develop on their own can be developed more quickly and efficiently when working with others. For example, observations of children engaged in peer interaction showed that children who had not yet reached the concrete operational stage could become conservers when working with children who had (Slavin et al., 2003). According to both Piaget and Vygotsky, when engaged in collaboration with peers, students debate one another, which causes them to expose their misconceptions and deliver better solutions (Slavin et al., 2003). While there is much theoretical and laboratory research on this perspective, there is a scarcity of classroom research that shows how cognitive developmental theory plays out in cooperative learning.

Cognitive elaboration perspective. The last perspective, cognitive elaboration theory, draws from cognitive psychology literature stating that in order to process new material, learners must perform cognitive restructuring (elaboration). Cognitive elaboration theorists believe that cooperative learning facilitates this process. Much evidence exists to suggest that explaining material to others is one of the most effective means of elaboration (Slavin et al., 2003). Research on cooperative learning also supports this perspective as students who gain the most from cooperative learning activities tend to be the students who provide the most elaborated explanations to others (Webb, 1989).

Based on the review of these four perspectives, it is clear that no single perspective is

sufficient to explain how cooperative learning functions. Instead, combining elements of all four perspectives reveals the necessary components that contribute to achievement during cooperative learning. First, combining group rewards and teaching students strategies that foster cooperative learning appears to produce better outcomes than either does alone. Second, cooperative learning appears to work more efficiently when there is a combination of group goals and individual accountability. Finally, all perspectives suggest that cooperative learning is effective because it makes private speech public, a factor that appears to help other group members better understand their thought processes behind problem solving. It seems, therefore, that though all students, regardless of academic competence, can benefit from cooperative learning, it is the high achievers who generally benefit the most. This result is possibly due to the fact that they give more elaborated explanations, which is linked with increased learning. Thus, cooperative learning can be an effective instructional strategy to increase student achievement, however its effectiveness depends on how it is implemented in the classroom. In the following section, I will discuss how status hierarchies in the classroom can influence the effectiveness of cooperative learning.

Status and Cooperative Learning

Cooperative learning is currently experiencing a renewal in popularity in schools (O'Donnell & Hmelo-Silver, 2013). This increased popularity is due to the fact that, unlike other types of instructional strategies, it accomplishes several academic and social goals such as increasing student achievement, motivation, self-esteem, liking for school, allowing for more opportunities for students to engage in classroom discourse, and teaching students social skills necessary for success in future schooling and the workplace (Antil, Jenkins, Wayne, & Vadasy, 1998; Emmer & Gerwels, 2002). It appears that when employed properly, cooperative learning

can be an invaluable tool to create more equitable student interactions in classrooms (Cohen & Lotan, 1995).

Some would argue, however, that like the classrooms in which they are situated, cooperative groups are vulnerable to the formation of status hierarchies (Garandean et al., 2013). *Status characteristic theory* (SCT) presents a mechanism for how status hierarchies occur in groups, thus allowing certain individuals to become more influential in a group. SCT suggests that during cooperative learning tasks, individuals use status characteristics to determine the competence of fellow group members, thereby creating a status hierarchy that affects other group members' access to participation (Melamed & Savage, 2013). Group members who are perceived as being higher in status contribute more, have a greater influence in group decisions, and are evaluated more favorably than group members of lower status (Melamed & Savage, 2013).

The formation of status hierarchies in cooperative groups decreases the effectiveness of cooperative learning by affecting the necessary components of successful cooperative learning identified in the theories presented above. First, having status hierarchies in a classroom can decrease social cohesion (Christianakis, 2010). In a classroom where the status hierarchy was based on race and academic competence, high status students only wanted to work with other high status students, black students only wanted to work with other black students, and white students only wanted to work with other white students, creating an in-group/out-group mentality (Christianakis, 2010). Second, the presence of status hierarchies can affect task motivation because high status students who are worried about grades typically take control of the group. In turn, this increases negative interdependence because high status students are denying low status students the opportunity to participate, thus preventing low status students from achieving their goal of learning, in order to protect their own goal of receiving a good grade (Christianakis,

2010).

Status hierarchies can also influence the quality of cognitive processing that typically occurs during cooperative learning. For instance, low status students, whose status was measured by popularity and perceived academic achievement, were ignored when they tried to share in cooperative groups, even when their contributions were valuable, thus denying the group a chance to learn from those students (Bianchini, 1997). Additionally, a necessary component in Vygotsky's theory is transfer of control, where less-able learners take agency of their learning in order to internalize the process. However, when engaged in a peer editing task, high status students tended to do the work for the low status students, resulting in a situation where low status students were not able to internalize the related cognitive processes (Christianakis, 2010). Status hierarchies also threaten intersubjectivity as the ideas of low status students are dismissed, and, therefore, neither student benefits from debating different ideas, exposing misconceptions, and arriving at a joint understanding (Hogan & Tudge, 1999).

According to SCT, status characteristics tend to be activated in groups when disagreements occur. Other group members use the status characteristics to determine whether to agree or disagree with their group members, even when the status characteristics do not affect other members' competence (Melamed & Savage, 2013). Qualitative analysis of videotape data suggests that group members sometimes use social cues such as eye contact, voice quality, and voice volume to determine when to ignore or reject proposals (Barron, 2003). Vygotsky too, referred to these traits that are unique to an individual as individual characteristics and acknowledged that they play a role in social interactions (Hogan & Tudge, 1999). Examples of individual characteristics include gender, physical attributes, and temperament. Personality is another individual characteristic that could be used to determine competence and, thereby,

influence one's status in groups. If status hierarchies are determined using traits that do not relate to actual competence, then cooperative learning can be compromised, as the most prominent voice may not be the most competent.

Psychological Type Theory

Psychological type theory is a method of classifying an individual's personality based on how they prefer to interact with their environment, process information, make decisions, and their general attitude towards life (Fourqurean, Meisgeier, & Swank, 1990). Carl Jung first defined psychological type theory in 1921 and published *Psychological Type* in 1923. According to Jung, seemingly random behavior actually follows specific patterns that tell us where an individual gets their energy, how they perceive information, and how they make decisions (Murphy & Meisgeier, 2008).

In adults, psychological type is often measured using the Myers-Briggs Type Indicator (MBTI) an instrument based on Jung's theory and developed by mother and daughter team Kathleen Briggs and Isabel Briggs Myers (Myers, McCaulley, Quenk & Hammer, 1998). The MBTI is the most popular personality instrument for use in non-psychiatric populations and is widely used by businesses and schools to classify individuals on four different dimensions of psychological type: 1. *Introversion (I)* and *Extraversion (E)* 2. *Sensing (S)* and *Intuition (N)*, 3. *Thinking (T)* and *Feeling (F)*, and 4. *Judging (J)* and *Perceiving (P)* (Murray, 1990). These four dimensions are further separated into *functions*, *attitudes*, and *orientations*. Sensing, Intuition, Thinking, and Feeling are attitudes and they form the basis of our personality. Each function is expressed with either an introverted or extraverted attitude. Finally, Judging and Perceiving are considered to be orientations (Murphy & Meisgeier, 2008).

Introversion and Extraversion. The Introversion and Extraversion dimension

determines where we focus our energy. Individuals with a preference for Introversion focus their energy on the inner world of ideas. Their best thinking occurs when they are alone. In contrast, individuals who prefer Extraversion gain their energy through interactions with people and events occurring in the outside world. They tend to work through their problems through talking.

Sensing and Intuition. The Sensing and Intuition dimension determines how we best take in information. Individuals who prefer Sensing use their senses to take in information from their environment, which they then compare and contrast with their existing knowledge. They learn best when presented with facts and concrete examples. Like individuals who have a preference for Sensing, individuals with a preference for Intuition also take in information using their senses, however instead of comparing and contrasting new information with existing knowledge, they look for patterns, connections, and meanings beyond what is immediately visible.

Thinking and Feeling. The Thinking and Feeling dimension tells us how we make decisions. A preference for Thinking results in logical and analytical decision making based on sound reasoning with the goal being to increase fairness and finding the objective truth. On the other hand, an individual with a preference for Feeling will take into consideration possible effects on people, relationships, and their own values.

Judging and Perceiving. The Judging and Perceiving orientation tells us how we prefer to organize the outer world. Individuals with a preference for Judging prefer order, planning, and structure, while individuals who prefer Perceiving favor a more spontaneous and flexible lifestyle.

Psychological Type Development

Jung maintained that the four functions (Sensing, Intuition, Thinking, and Feeling)

develop over time (Murphy & Meisgeier, 2008). Each individual has a dominant function, which is the first to develop and which is expressed with either an introverted or extraverted attitude. A child's dominant function is still developing during their elementary school years and it is believed that their dominant function is not fully developed until the age of 14 (Murphy & Meisgeier, 2008). Individuals also have auxiliary, tertiary, and inferior functions, which develop throughout their lifetime. The developmental stages of these functions vary by individual, with some having their auxiliary function fully developed by the age of 20 but others taking longer (Myers & Kirby, 2012). This developmental view of psychological type is supported by the research, at least for the Thinking and Feeling dimension, as more elementary school students prefer Feeling and more secondary school students prefer Thinking (Sak, 2004).

Psychological Type as a Status Characteristic

Research on psychological type in higher education and the workplace suggests that personality traits influence communication and performance in cooperative learning (Opt & Loffredo, 2000; Ostafichuk & Naylor, 2013). In terms of communication, college students with a preference for Introversion and Sensing scored higher for communication apprehension, or anxiety of communicating with other people whether in groups, dyads, at a meeting, or public speaking, than students who preferred Extraversion and Intuition. (Opt & Loffredo, 2000). Additionally, when using dominance as an indicator of communicator image, extraverts have been shown to see themselves as more competent communicators than introverts (Opt & Loffredo, 2003). Not surprisingly, people with a higher dominant trait tend to have a greater influence when working in groups because they are rated as more competent by group members, even when controlling for actual abilities (Anderson & Kilduff, 2009).

Psychological type has also been shown to influence the success of the group in

cooperative learning (Ostafichuk & Naylor, 2013). For instance, research suggests that extraverts have a greater influence in groups even when they are not necessarily the most competent group member. Groups are more likely to make an incorrect decision when a student with a preference for Extraversion also had that incorrect answer and are less likely to accept the correct answers given by students with a preference for Introversion, and this influence becomes stronger as time progresses (Ostafichuk & Naylor, 2013). Interestingly, this effect was observed in an environment where students had attended workshops about MBTI and had learned about the implications for group work.

Finally, a study of junior high students suggests that personality can influence classroom status (van der Linden, Scholte, Cillessen, Nijenhuis, & Seegers, 2010). Students took the Big Five personality test and were then asked to nominate peers who were most liked and who were most popular. The results showed that Extraversion was significantly correlated with likeability, the extent to which one is considered friendly and cooperative, and popularity, the extent to which one has influence in a group (van der Linden et al., 2010).

It is likely that psychological type interacts with other status characteristics to create status hierarchies in cooperative learning groups, where certain types are perceived as more competent than others, thus preventing individuals with certain psychological types from participating and displaying their competence. Additionally, because research has demonstrated the negative effects of status hierarchies for both low-status individuals and the success of the overall group, exploring the possibility of psychological type as another status characteristic is worthy of investigation. There are few studies that explicitly measure the relationship between psychological type and social status in the classroom, and fewer studies still that assess the psychometric properties of instruments used to determine psychological type in children and

adolescents. The following review will begin to fill this gap in the literature. In this next section, I will investigate the psychometric properties of two instruments used to determine psychological type in children and adolescents.

Measuring Psychological Type in Children and Adolescents

Murphy-Meisgeier Type Indicator for Children. In children, psychological type is measured in several ways including using the MBTI and ethnographically, through researcher observation techniques. However, difficulties have occurred because students have reported that the MBTI is too difficult for middle schoolers to understand and adult observations are unreliable (Gilbert, 1998; Murphy & Meisgeier, 2008). To resolve these challenges, researchers developed the Murphy-Meisgeier Type Indicator for Children (MMTIC), which is the first instrument for children that measures the same constructs as the MBTI. Developed in 1987, the MMTIC's purpose was to assess students' psychological type while they were still in school (Murphy & Meisgeier, 2008). As time passed, a revised version of the instrument was released in 2008 to reflect advances made regarding how psychological type develops in children. This revised MMTIC (2008) is a 43-item self-report instrument used to measure the psychological type of children in grades 2 through 12. It is available both online and in pencil and paper form. Both versions can be administered in a group setting in about 15 to 20 minutes. Each question asks the participants to indicate their preference between two choices. The questionnaire determines the students' preferences in the following four domains: Introversion (I) – Extraversion (E), Sensing (S) – Intuition (N), Feeling (F) – Thinking (T), and Judging (J) – Perceiving (P) The questionnaire results in a 4-letter indication of one of sixteen psychological types (INFJ, ESTP, and so on).

The developers calculated the psychometric properties of the revised instrument and

reported that Cronbach's alpha, an estimate of internal consistency, ranged from .61 (SN and JP) to .71 (TF) with an average of .64. Test-retest reliability ranged from .69 (JP) to .78 (EI) with an average of .73, and split-half reliability ranged from .57 (JP) to .69 (TF), with an average of .63 (Murphy & Meisgeier, 2008).

Student Styles Questionnaire. The Student Styles Questionnaire (SSQ) is another instrument used to measure psychological type in children and adolescents. It was developed in 1996 and is a 69-item self-report instrument meant for children ages 8 through 17. Like the MMTIC, it is based on Jung's psychological type theory with some adjustments to reflect a later time and circumstance. The Sensing-Intuition scale is replaced with Practical-Imaginative, and the Judging-Perceiving scale is replaced with Organized-Flexible. In this review, Sensing-Intuition (SN) and Judging-Perceiving (JP) will be used for clarity. The developers reported internal consistency to be in the high .90s and test-retest reliability coefficients ranged from .67 (SN) to .80 (EI), with a mean of .74 (Oakland, Glutting, & Horton, 1996).

Overall, the reliability estimates given for both the MMTIC and the SSQ do not consistently reach levels generally accepted in research. Therefore, it is necessary to review research using these two instruments to further understand their reliability. Confidence in the reliability of these instruments will help ensure that the results of studies that use these instruments to measure psychological type in children and adolescents are not compromised.

Reliability

In the next section I will examine reliability and its implications for consistency and accuracy in instruments used to measure psychological type in children and adolescents.

Sources of Error in Measurement. A certain amount of error exists when using any instrument to measure a phenomenon (Carmines & Zeller, 1979; Overton, 2011). Two types of

errors have been identified: *systematic* or *nonrandom* and *unsystematic* or *random*. Systematic errors are factors that affect the accuracy of instruments in a consistent manner, for example, a scale that consistently adds five pounds to a measurement. On the other hand, random errors are those that affect the measurement of a phenomenon randomly (Carmines & Zeller, 1979; Drost, 2011). There are too many sources of random error to list every one, however here are a few common examples. First, there are factors that affect the subject being measured such as fatigue or hunger. Another source of random error occurs when the observer, rater, or interviewer misunderstands or misinterprets the subject's response. A third source of random error is the environment where the instrument is being used, such as the presence of a loud noise in the middle of testing. The instrument itself can be a source of random error depending on word choice and questions asked. Finally, random error can occur during data processing when data is being entered or analyzed (Weiner, 2007). Because researchers are aware that random error exists in all instruments, they must determine the extent to which an instrument contains random error so that instruments with the smallest amount of error can be used to collect data (Carmines & Zeller, 1979; Overton, 2011).

Classical Test Theory. *Classical Test Theory*, (also called Classical test score theory, or test theory), is a model used to assess random measurement error (Carmines & Zeller, 1979). This theory is attributed to Charles Spearman and was updated and modernized by Lord and Novick in 1968 (Junker, 2012). The basic premise behind the theory is that we must consider random error in our results, meaning that the observed score on an instrument actually consists of both a true score and an error score (Carmines & Zeller, 1979; Junker, 2012). A true score is the average score that would be obtained if an instrument were administered an infinite number of times, while the error score is the difference between the observed score and the true score

(Carmines & Zeller, 1979; Robinson, 2008).

Because of the presence of random error, each administration of an instrument will result in a different observed score. Due to the laws of probability, it is just as likely for random error to cause an increase in the observed score, as it is to cause a decrease. Therefore, it is assumed that while the error will always be present in the observed score, multiple administrations of an instrument will cause the observed score to equal out and be consistent (Carmines & Zeller, 1979). This tendency towards consistency is referred to as *reliability* (Carmines & Zeller, 1979). Thus, the more consistent the observed score over multiple administrations of the instrument, the higher the reliability, which indicates that less random error is present in the observed score (Carmines & Zeller, 1979). The smaller amount of random error there is, the more confidence we can have that the instrument is consistently measuring the phenomenon (Gay, Mills, & Airasian, 2011). Determining reliability is important but certain situations may make administering an instrument multiple times unfeasible or undesirable. Therefore, researchers use several different methods to measure reliability that will be discussed in further detail later on in this paper.

Pearson's r. Reliability is concerned with whether an instrument reports similar results when administered to the same group of people repeatedly (Carmines & Zeller, 1979; Drost, 2011; Gay et al., 2011; Junker, 2012; Oluwatayo, 2012; Overton, 2011). It is important to note that reliability is not just a property of an instrument, but also a property of the scores (Junker, 2012). Reliability is expressed numerically using a *reliability coefficient* (Drost, 2011; Gay et al., 2011). A common reliability coefficient is Pearson's r , (attributed to Karl Pearson, 1867 – 1936), also called Pearson's Product Moment correlation (Overton, 2011). It is used with interval or ratio data and describes the strength of a relationship between two variables (Klockers &

Sanders, 2012). The coefficient can have a value between -1 and 1 (Overton, 2011). A test with perfect reliability would have a coefficient of 1 or -1 (Gay et al., 2011). Generally researchers state that a score of .60 or greater indicates adequate reliability but a score of .80 or greater is preferred (Gilbert, 1998; Overton, 2011). A reliability coefficient also expresses how much variance in an observed score can be attributed to variance in the true score versus measurement error (Webb, Shavelson, & Haertel, 2006). A high reliability coefficient indicates low measurement error (Gay et al., 2011). For example, a reliability coefficient of .80 is interpreted as meaning that 80% of the variance is attributable to true differences (true score) while 20% is attributable to measurement error.

Researchers are concerned with determining three types of consistency when approaching reliability of instruments: consistency over time, consistency of items on an instrument, and equivalence (Drost, 2011; Overton, 2011). Thus, different tests are used to determine the different types of consistency. In the following section, I will give a brief overview of several methods researchers use to determine reliability.

Test-Retest Reliability. Consistency over time is also called *stability* (Drost, 2011; Gay et al., 2011). If an instrument is given repeatedly and the scores are similar on each administration, the instrument is considered to have high stability (Gay et al., 2011). Having high stability is important for instruments that are used to make predictions because the instrument's predictive quality depends on whether the scores will remain stable over time (Gay et al., 2011). Test stability is estimated using *test-retest reliability*, where an instrument is administered twice to the same group of people and the two sets of scores are correlated using Pearson's r (Carmines & Zeller, 1979; Gay et al., 2011). A correlation coefficient of .80 or greater indicates acceptable test-retest reliability (Oluwatayo, 2012).

While test-retest reliability is good for determining stability of a test over time, there are, however, several limitations to this methodology. First, the process of measuring the phenomenon may cause the phenomenon to change, which is called *reactivity* (Carmines & Zeller, 1979). For instance, after taking the test the first time, participants may research the phenomenon being studied, which may cause them to change their responses on the second administration. The duration of time between administrations can also affect test-retest reliability. If the time between administrations is too small, participants may remember their previous responses. On the other hand, if too much time has passed between tests then maturation, changes within the participants that occur over time, may affect the results (Carmines & Zeller, 1979; Drost, 2011; Oluwatayo, 2012; Overton, 2011).

Internal Consistency Reliability. Another way in which reliability is estimated is by looking at the extent to which items on a test are consistent among themselves and the test as a whole (Gay et al., 2011). This is referred to as *internal consistency*, which determines the extent to which items on a test or subscale are measuring the same construct (Tavakol & Dennick, 2011). Internal consistency has an advantage over test-retest reliability because it only requires one administration of a test. This eliminates the effects of maturation and reactivity and also reduces other errors that may occur as a result of differences in testing conditions (Gay et al., 2011). Internal consistency is measured using *split-half reliability*, *Kuder-Richardson*, or *Cronbach's alpha*. Below I will discuss split-half reliability and Cronbach's alpha.

Split-Half Reliability. Split-half reliability is a measure of internal consistency where an instrument is divided into two halves and the scores on those two halves are correlated using Pearson's r (Gay et al. 2011; Carmines & Zeller, 1979; Overton, 2011). If an instrument has high internal consistency, it is expected that scores on one half will correlate highly with scores on the

other half (Oluwatayo, 2012). Like test-retest reliability, a coefficient of .80 or greater indicates an acceptable level of reliability (Oluwatayo, 2012).

To estimate split-half reliability, an instrument is administered, then it is divided into two halves and the participants' scores on the two halves are computed and correlated using Pearson's r , then are corrected for test length using the Spearman-Brown formula (Gay et al., 2011; Oluwatayo, 2012). As stated earlier an advantage of using split-half reliability is that it eliminates the complications involved with administering a test multiple times. However, a significant disadvantage is that the results depend on how the test is split in half (Drost, 2011). A common method of splitting an instrument is by assigning odd test items to one half and even test items to the other but even this method can pose problems for researchers (Gay et al., 2011; Oluwatayo, 2012).

Cronbach's alpha. A second measure of internal consistency is Cronbach's alpha. Cronbach's alpha was developed by Lee Cronbach in 1951 to provide a measure of internal consistency of a test or a scale (Tavakol & Dennick, 2011). It is expressed as a number between 0 and 1. A score above .70 is considered an acceptable measure of reliability (Oluwatayo, 2012; Tavakol & Dennick, 2011). Cronbach's alpha is the most popular measure of internal consistency reliability (Carmines & Zeller, 1979; Drost, 2011). It has an advantage over both test-retest and split-half reliability as it does not require splitting test items and can be calculated after one administration (Carmines & Zeller, 1979).

Present Study

Researchers prefer to choose instruments with high reliability, as they can be more confident when interpreting their results. Both the MMTIC and the SSQ show questionable reliability as reported in the test manual. The purpose of the present study is to conduct a

systematic review on the empirically derived reliability of the MMTIC and the SSQ, two instruments used to measure psychological type in children and adolescents. Reviewing the sample-specific reliability of these instruments when used in research studies will help researchers determine which instrument is most appropriate for a proposed study. Choosing reliable instruments will ensure that the results of their studies are not compromised so that researchers can continue to explore the effect of psychological type in education, particularly on its role in determining classroom status for child and adolescent learners.

Method

Search Procedure

A search for studies that used the MMTIC or SSQ to measure psychological type in children and adolescents was conducted in the following databases: 1) Academic Search Complete, 2) Educational Source, 3) ERIC, 4) Health and Psychosocial Instruments, 5) Teacher Reference Center, 6) PsycINFO, 7) PsycArticles, 8) Primary Search, and 9) ProQuest Dissertations & Theses.

The following search terms were used 1) (“psychological type” AND (adolescen* or child*) NOT (MBTI OR Myers-Brigg?)), 2) (MMTIC OR “murphy-meisgeier type indicator for children” OR “murphy-meisgeier type indicator” OR “murphy meisgeier”), and 3) (“Student Styles Questionnaire”). Hand searching was done in *The Journal of Psychological Type*. All articles were cross-referenced. The searches were carried out in May 2014.

After duplicate studies were removed, the study titles and abstracts were screened for relevance. Inclusion criteria were: 1) studies that used the SSQ or MMTIC, 2) English speaking participants using the English version of the instruments, and 3) School aged participants. There was no restriction on participant gender or race. A total of 146 articles meeting the criteria were

identified.

The full text of the remaining 146 articles was examined to determine whether the authors reported at least one reliability estimate. Of the 146 articles, only 6 (4%) included at least one reliability estimate. The reliability estimates from the 6 articles were extracted and compared. This resulted in reliability estimates from 4 articles that could be compared for test-retest reliability and internal consistency reliability.

Results

In this next section, I will outline the results that emerged from a review of the reliability of instruments that measure psychological type in children and adolescents.

General Study Characteristics

The four studies differed in terms of number of participants, participant demographics, and instrument used (see Table 1). The number of participants ranged from 81 (McPeek & Breiner, 2013) to 7902 (Callueng, 2012). Participant ages ranged from 8 (Callueng, 2012; Oakland, Glutting, & Horton, 1996) to 19 years old (McPeek & Breiner, 2013). One study (Gilbert, 1998) reported grade level instead of age. The proportion of male to female participants ranged from 26/74 (McPeek & Breiner, 2013) to 50/50 (Callueng, 2012). Finally, two studies used the SSQ (Callueng, 2012; Oakland, Glutting, & Horton, 1996), one study used the original MMTIC (Gilbert, 1998), and one study used the revised MMTIC (McPeek & Breiner, 2013).

Table 1

<i>Study Characteristics</i>							
Study	<i>n</i>	Male	%	Female	%	Age range (years)	Instrument
Callueng (2012)	7902	3951	50	3951	50	8-17	SSQ
Gilbert (1998)	252	131	52	121	48	14-18 ^a	MMTIC (1987)
McPeek & Breiner (2013)	81	21	26	60	74	15-19	MMTIC (2008)
Oakland, Glutting, & Horton (1996)	137	75	55	62	45	8-17	SSQ

^a Ages were estimated as study only reported participant grades (8th-12th)

Test-Retest Reliability

The amount of time that passed between administrations of the instruments differed significantly between the studies with the shortest amount of time being two weeks (McPeek & Breiner, 2013) and the longest amount of time being two years (Gilbert, 1998). Overall, only the revised MMTIC reached an acceptable level of test-retest reliability with a mean coefficient of .81 for all scales (see Table 2). Both the EI and JP scales reported the highest reliability with a coefficient of .83. The TF scale reported the lowest reliability with a coefficient of .77 and was the only scale that did not reach an acceptable level of test-retest reliability (McPeek & Breiner, 2013). The SSQ reached an acceptable level of reliability for the EI scale (.80) but not for any of the other scales. The lowest reported coefficient was .67 for the SN scale (Oakland, Glutting, & Horton, 1996). The original MMTIC reported the lowest coefficients with a mean of .57 for all four scales. None of the scales reached an acceptable coefficient of .80 (Gilbert, 1998).

Table 2

Test-Retest Reliability

Study	Instrument	Time between Administrations	EI	SN	TF	JP	Mean
Gilbert (1998)	MMTIC (1987)	Two years	.52	.55	.59	.61	.57
McPeek & Breiner (2013)	MMTIC (2008)	Two weeks	.83	.82	.77	.83	.81
Oakland et al. (1996)	SSQ	Seven months	.80	.67	.78	.74	.75

Split-Half Reliability

Overall, the revised MMTIC reported the strongest split-half reliability with an average coefficient of .76 for all scales (see Table 3). The highest scales were EI and TF (.78) and the lowest scale was JP (.73). None of the scales reached an acceptable coefficient level of .80

(McPeck & Breiner, 2013). In comparison, the JP scale on the original MMTIC reported the highest coefficient (.82), and was the only scale that reached the acceptable level. The other scales reported lower coefficients with a mean coefficient of .75, and the SN scale showed the weakest reliability (.71) (Gilbert, 1998).

Table 3

<i>Split-Half Reliability</i>						
Study	Instrument	EI	SN	TF	JP	Mean
Gilbert (1998)	MMTIC (1987)	.72	.71	.75	.82	.75
McPeck & Breiner (2013)	MMTIC (2008)	.78	.75	.78	.73	.76

Cronbach's alpha

On another measure of internal consistency, the revised MMTIC yielded the strongest mean coefficient alpha of .76 across all scales, which also matched the split-half reliability estimate (see Table 4). The JP scale reported the lowest coefficient alpha (.74), however all scales reached the acceptable alpha level of .70 (McPeck & Breiner, 2013). In contrast, none of the scales on the SSQ reached the acceptable level of .70. The alpha coefficients ranged from .35 for the TF scale to .69 for the EI scale, with a mean alpha coefficient of .57 for all of the scales (Callueng, 2012).

Table 4

<i>Cronbach's alpha</i>						
Study	Instrument	EI	SN	TF	JP	Mean
Callueng (2012)	SSQ	.69	.55	.35	.68	.57
McPeck & Breiner (2013)	MMTIC (2008)	.75	.77	.78	.74	.76

Discussion

Given that research on psychometric properties of instruments used to measure psychological type in children and adolescents is almost non-existent, I conducted a systematic literature review of studies that used two instruments in particular, the Murphy-Meisgeier Type Indicator for Children (MMTIC) and the Student Styles Questionnaire (SSQ). Reliability estimates were extracted and compared in order to determine which instruments provided adequate reliability. Results indicated that the revised MMTIC (2008) yielded stronger internal consistency and test-retest reliability estimates, although differences in reliability existed between scales.

Based on the data reported, the revised MMTIC displayed the better test-retest reliability and internal consistency reliability. The revised MMTIC showed the best stability, meaning that the scores obtained during the first administration were highly correlated with the scores obtained during the second administration. The TF scale was the one scale that did not reach an acceptable level of reliability. Previous studies have shown that the TF dimension is heavily affected by development, with older participants displaying a preference for Thinking while younger participants display a preference for Feeling (Murphy & Meisgeier, 2008; Sak, 2004). However the results from this review do not support this finding. First, only two weeks separated administrations, which is likely to be too short for maturation effects to occur. Second, this finding was not seen in Gilbert (1998), which had a much longer time period between administrations. Therefore, it appears that the TF scale of the revised MMTIC is less stable over time, compared to the other scales.

Compared to the original MMTIC and the SSQ, the revised MMTIC also reported stronger internal consistency, meaning that the items on each subscale and the instrument as a

whole were strongly correlated with each other. A surprising finding was that the SSQ showed much lower alpha coefficients compared to the reliability estimates reported by the developers (Oakland et al., 1996). Low internal consistency indicates that the test items may not be assessing the same construct. This would have serious implications for research studies.

Participant age is one factor that could have influenced the reliability differences found between the two instruments as the ages of the participants varied across the studies. In general, younger children tend to be less consistent in their reporting than older children and adults (Murphy & Meisgeier, 2008). Thus, age could account for some of the differences found between the reliability coefficients of the revised MMTIC reported by McPeck and Breiner (2013) who's youngest participants were 15 years old, and the other studies who had participants as young as 8 years old, including the test developers (Murphy & Meisgeier, 2008).

Limitations

There are several limitations to this review. First, only a small number of studies were included because they were the only studies that met the criteria of reporting at least one reliability coefficient. The small N limits what we can generalize about the instruments. A related limitation is that the studies did not report the same reliability coefficients so not all of the instruments could be compared on the same estimates of reliability. A third limitation is that the decisions made by the researchers of the original studies affected the quality of the studies included in this review. For example, McPeck and Breiner (2013) only included 81 participants. Additionally, Gilbert (1998) waited two years between administrations of the instrument, instead of the recommended time period of two weeks to a month, which may have affected the test-retest reliability results (Carmines & Zeller, 1979). A final limitation is that only one researcher selected and coded the studies, meaning some relevant studies may have been excluded and

inappropriate studies could have been included in the analysis.

Future Research

Studying psychological type in children and adolescents, particularly as it relates to education would provide benefits for researchers, teachers, and students. However, as this study shows, instruments used to measure psychological type in children and adolescents must continue to be developed and refined in order to improve reliability of both the overall instrument and of certain scales. Additionally, because the instruments, particularly the SSQ, displayed low internal consistency, which suggests a deficit in the construct being measured, future studies should also investigate the validity of these instruments. Once the psychometric properties of these instruments are established, researchers can and should use these instruments to measure psychological type in children and adolescents to help inform educational practices, such as status issues in the classroom.

Implications

Based on the results of this study, the revised MMTIC provides the higher level of test-retest and internal consistency reliability among instruments used to measure psychological type in children and adolescents. However, researchers should still use caution when using this instrument, as certain scales reported low levels of reliability. For example, the TF scale did not reach an appropriate test-retest reliability coefficient level meaning scores that a participant achieves on the first administration may differ when tested again in the future.

An unintended consequence of this study was revealing the large number of researchers who used these instruments but failed to calculate and report reliability estimates in their studies, which demands the question as to why these researchers chose not to investigate reliability. One possibility is that because they were working with students in classrooms, they were limited in

how much time they had and were unable to conduct repeated assessments, making test-retest reliability difficult. If this were the case, they could still calculate internal consistency. Another possibility is that the researchers were not concerned with whether the scores would remain consistent over time because they were investigating the effects of the participants' present psychological type on another variable. A third option is that they were content with the reliability estimates reported by the test developers in the manual. Regardless, an instrument's reliability is population dependent, meaning a reliability coefficient indicates whether an instrument can be considered reliable for a certain population under certain conditions, thus reliability should be considered when using any instrument in a new setting or with a new population (Junker, 2012; Weiner, 2007).

In summary, a review of previous research suggests that psychological type may act as a status characteristic in the classroom and contribute to the creation of status hierarchies. This influences how students interact with one another when participating in cooperative learning, and in addition, tends to negatively affect student social-emotional wellbeing. This study suggests that gaining a greater understanding of how psychological type in children and adolescents affects participation in cooperative learning has great potential to help teachers create more equitable classrooms.

References

- Anderson, C., & Kilduff, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *Journal of Personality and Social Psychology, 96*, 491 – 503.
- Antil, L. R., Jenkins, J. R., Wayne, S. K., & Vadasy, P. F. (1998). Cooperative learning: Prevalence, conceptualizations, and the relation between research and practice. *American Educational Research Journal, 35*, 419 – 454.
- Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*, 307-359.
- Bianchini, B. (1997). Where knowledge construction, equity, and context intersect: Student learning of science in small groups. *Journal of Research in Science Teaching, 34*, 1039 – 1065.
- Callueng, C. M. (2012). *Cross-national study of children's temperament: Structural validity of the student styles questionnaire* (Unpublished doctoral dissertation). University of Florida, Gainesville, FL.
- Capraro, R. M. & Capraro, M. M. (2002). Myers-Briggs Type Indicator score reliability across: Studies a meta-analytic reliability generalization study. *Educational and Psychological Measurement, 62*, 590 – 602.
- Carmines, E. G. & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage.
- Christianakis, M. (2010). “I don't need your help!” Peer status, race, and gender during peer writing interactions. *Journal of Literacy Research, 42*, 418 – 458.
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1 – 35.
- Cohen, E. G. & Lotan, R. A. (1995). Producing equal-status interaction in the heterogenous classroom. *American Educational Research Journal, 32* (1), 99 – 120.
- Dembo, M. & McAuliffe, T. (1987). Effects of perceived ability and grade status on social interaction and influence in cooperative groups. *Journal of Educational Psychology, 79*, 415 – 423.
- Drost, E. A. (2011). Validity and reliability in social science research. *Education Research and Perspectives, 38*, 105 – 123.
- Emmer, E. T. & Gerwels, M. C. (2002). Cooperative learning in elementary classrooms: Teaching practices and lesson characteristics. *The Elementary School Journal, 103* (1), 75 – 91.
- Fourqorean, J. M., Meisgeier, C., & Swank, P. (1990). The link between learning style and Jungian psychological type: A finding of two bipolar preference dimensions. *The Journal of Experimental Education, 58* (3), 225 – 237.
- Furnham, A. (1996). The big five versus the big four: The relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five-factor model of personality. *Journal of Personality and Individual Differences, 21*, 303 – 307.
- Garandau, C. F., Lee, I. A., & Salmivalli, C. (2013). Inequality matters: Classroom status hierarchy and adolescents' bullying. *Journal of Youth and Adolescence*. Advance online publication. doi: 10.1007/s10964-013-0040-4
- Gay, L. R., Mills, G. E., & Airasian, P. W. (2011). *Educational research: Competence for analysis and applications*. Pearson Education.
- Gilbert, A. P. (1998). *A test-retest study of the Myers-Briggs Type Indicator and the Murphy-Meisgeier Type Indicator for Children over a two year time period* (Doctoral dissertation).

- Retrieved from ProQuest Dissertations and Theses database (UMI No. 9833981).
- Hogan, D. M. & Tudge, J. R. H. (1999). Implication of Vygotsky's theory for peer learning. In O'Donnell, A. M. & King, A. (Eds.) *Cognitive Perspectives on Peer Learning* (39 – 65). Mahwah, NJ.
- Johnson, D. W. & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, 38, 365 – 379.
- Jonkmann, K., Trautwein, U., & Ludtke, O. (2009). Social dominance in adolescence: The moderating role of classroom context and behavior heterogeneity. *Child Development*, 80 (2), 338 – 355.
- Junker, B. W. (2012). *Some aspects of classical reliability theory & classical test theory* [PDF document]. Retrieved from <http://www.stat.cmu.edu/~brian/PIER-methods/For%202013-02-26/Readings/CTT-formulas.pdf>
- Katz, I. & Benjamin, L. (1960). Effects of white authoritarianism in biracial work groups. *Journal of Abnormal Social Psychology*, 5, 448 – 556.
- Klockers, A. J. & Sanders, E. A. (2012). *EDPSY 490: Basic educational statistics*. Seattle, WA: University of Washington.
- Linton, R. (1936). *The study of man: An introduction*. New York: Appleton-Century Company, Inc.
- McPeck, R. W. & Breiner, J. (2013). Convergent validity of the MBTI and MMTIC assessments. *Journal of Psychological Type*, 73, 7 – 20.
- Melamed, D. & Savage, S.V. (2013). Status, numbers and influence. *Social Forces*, 91, 1085-1104.
- Modin, B., Ostberg, V., & Almquist, Y. (2011). Childhood peer status and adult susceptibility to anxiety and depression. A 30-year hospital follow-up. *Journal of Abnormal Child Psychology*, 39, 187 – 199.
- Morine-Dershimer, G. (1983). Instructional strategy and the “creation” of classroom status. *American Educational Research Journal*, 20(4), 645 – 661.
- Murphy, E. & Meisgeier, C. (2008). *MMTIC manual: A guide to the development and use of the Murphy-Meisgeier type indicator for children*. Center for Applications of Psychological Type.
- Myers, K. D. & Kirby, L. K. (2012). *Introduction to type dynamics and development*. CPP, Inc.
- Myers, I. B., McCaulley, M. H., Quenk, N. L., & Hammer, A. L. (1998). *MBTI manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Mountain View, CA: CPP, Inc.
- Oakland, T., Glutting, J. J., & Horton, C. B. (1996). *Student Styles Questionnaire: star qualities in learning, relating, and working*. San Antonio, TX: The Psychological Corporation.
- O'Donnell, A. M. & Hmelo-Silver, C. E. (2013). Introduction: What is collaborative learning?: An overview. In C.E. Hmelo-Silver, C.A. Chinn, C.K.K. Chan, & A. O'Donnell (Eds.), *The International Handbook of Collaborative Learning* (1 – 15). New York: Routledge.
- Oluwatayo, J. A. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research*, 2, 391 – 400.
- Opt, S. K. & Loffredo, D. A. (2000). Rethinking communication apprehension: A Myers-Briggs perspective. *The Journal of Psychology*, 134 (5), 556 – 570.
- Ostafichuk, P. M. & Naylor, C. (2013, June). *The influence of personality type on teamwork in engineering education*. Paper presented at the Canadian Engineering Education Association Conference, Montreal, Canada.

- Overton, T. (2011). *Assessing learners with special needs: An applied approach*. Upper Saddle River, NJ: Pearson Education.
- Robinson, A. (2008, September). What is reliability in terms of classical test theory? [Web log comment]. Retrieved from http://surveymethodsaddicts.blogspot.com/2008/09/what-is-reliability_28.html
- Rosenholtz, S. J. (1985) Treating problems of academic status. In J. Berger & M. Zelditch, Jr. (Eds.) *Status, rewards, and influence* (pp. 445 – 470). San Francisco: Jossey-Bass.
- Rosenholtz, S. J. & Wilson, B. (1980). The effects of classroom structure on shared perceptions of ability. *American Educational Research Journal*, *17*, 175 – 182.
- Sak, U. (2004). A synthesis of research on psychological types of gifted adolescents. *Journal of Secondary Gifted Education*, *15*, 70 – 79.
- Slavin, R. E., Hurley, E. A., & Chamberlain, A. (2003). Cooperative learning and achievement: Theory and research. In Reynolds, W. M., Miller, G. E. & Weiner, I. B. (Eds.) *Handbook of psychology: Vol. 7. Educational psychology* (pp. 177 – 198). New Jersey: John Wiley & Sons.
- Tavakol, M. & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53 – 55.
- Van der Linden, D., Scholte, R. H. J., Cillessen, A. H. N., Nijenhuis, J. T., & Segers, E. (2010). Classroom ratings of likeability and popularity are related to the Big Five and the general factor of personality. *Journal of Research in Personality*, *44*, 669 – 672.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.)). Cambridge: Massachusetts: Harvard University Press.
- Webb, N. M. (1984). Sex differences in interaction and achievement in cooperative small groups. *Journal of Educational Psychology*, *76*, 33 – 44.
- Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, *13*, 21–39.
- Webb, N. M., Farivar, S. H., & Mastergeorge, A. M. (2001). *Productive helping in cooperative groups* (CSE Technical Report No. 555). Los Angeles: University of California.
- Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). Reliability coefficients and generalizability theory. *Handbook of Statistics*, *26*, 1 – 44.
- Weiner, J. (2007). *Measurement: Reliability and validity measures* [PDF document]. Retrieved from http://ocw.jhsph.edu/courses/hsre/PDFs/HSRE_lect7_weiner.pdf
- Wilson, T., Karimpour, R., & Rodkin, P. C. (2011). African American and European American students' peer groups during early adolescence: Structure, status, and academic achievement. *The Journal of Early Adolescence*, *31*, 74 – 98.