

THE EFFECTS OF FREE PLAY AS AN INSTRUCTIONAL TOOL ON THE QUALITY
OF IMPROVISATION OF FIRST, SECOND, AND THIRD GRADE CHILDREN

Tammie L. Burger, B.M, M.M.

Dissertation Prepared for the Degree of
DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2000

APPROVED:

William V. May, Major Professor
Hildegard Froehlich, Minor Professor
Darhyl Ramsey, Committee Member
Debbie Rohwer, Committee Member
Thomas S. Clark, Dean of the College of Music
C. Neal Tate, Dean of the Robert B. Toulouse School of
Graduate Studies

Burger, Tammie L., The effects of free play as an instructional tool on the quality of improvisation of first, second, and third grade children. Doctor of Philosophy (Music Education), December 2000, 161 pp., 17 tables, 8 figures, references, 62 titles.

To look at the effect of free play on the musical improvisations of first, second and third grade children, 108 children were randomly assigned to either a control or treatment group. Subjects were tested using a researcher-designed instrument to elicit an improvisatory response. The control group then received regular music instruction (120 minutes every 2 weeks) and the treatment group received regular music instruction in conjunction with musical free play (100 minutes of instruction and 20 minutes of free play every 2 weeks). The treatment lasted 14 weeks. At the end of the treatment, all students were tested with the same testing instrument used for the pre test. Videotapes of the improvisations were submitted to three independent judges to rate for quality on a 5-point Likert scale. The change in ratings between pre and post tests were analyzed with an analysis of variance to determine if there were significant differences between the control and treatment groups. The analysis of the data revealed no significant difference in the change of ratings between control and treatment groups for the group as a whole, or for any particular grade level within the total group.

TABLE OF CONTENTS

| | Page |
|---|------|
| LIST OF TABLES..... | v |
| LIST OF ILLUSTRATIONS..... | vi |
| Chapter | |
| 1. RATIONALE, PURPOSE, AND RESEARCH QUESTIONS..... | 1 |
| Rationale | |
| The Role of Play in Development | |
| The Role of Play in Musical Development | |
| Purpose and Research Questions | |
| Definition of Terms | |
| 2. REVIEW OF LITERATURE..... | 11 |
| Introduction | |
| Play Theory | |
| Early Play Theory | |
| Recent Play Theory | |
| Defining Play | |
| Play and Cognitive Development | |
| Play and Music | |
| Play and Music: Theoretical Considerations | |
| Quantification of the Relationship between Play and Music | |
| Issues of Measurement | |
| Consensual Assessment | |
| Consensual Assessment Applied to Musical Assessment | |

| | |
|--|-----|
| 3. METHODOLOGY..... | 56 |
| Introduction | |
| Design of the Study | |
| Threats to Research Design | |
| Statistical Considerations—Sample Size | |
| Research Setting | |
| Subjects | |
| Method | |
| Development of the Improvisation Task | |
| Improvisation Task | |
| Field Test I | |
| Development of the Free Play Protocol | |
| Field Test II—Free Play Protocol | |
| Main Study—Free Play Protocol | |
| Refining the Testing Procedure | |
| Field Test II—Testing Procedure | |
| Main Study—Testing Procedure | |
| Rating of Improvisations | |
| Consensual Assessment Technique | |
| Reliability Estimates | |
| Pilot Testing | |
| Testing Protocol for the Main Study | |
| Rating of Improvisations | |
| Rating Protocol for this Study | |
| 4. RESULTS..... | 85 |
| Descriptive Data | |
| Determination of Statistical Significance | |
| Homogeneity of Variance | |
| Analysis of Variance | |
| 5. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER STUDY..... | 102 |
| Summary | |
| Findings | |
| Research Design and Subsequent Findings | |
| Sample Size and Variability of the Measurement Instrument | |
| Interjudge Reliability | |
| Length of Study | |

| | |
|--|-----|
| Conclusions | |
| Play and Cognitive Development | |
| Treatment Considerations | |
| Task Considerations | |
| Conclusions | |
| Play and Music | |
| Issues of Measurement | |
| Recommendations for Further Study | |
| Personal Observations | |
| Suggestions for Implementation of Musical Play | |
| Personal Observations on Musical Play | |
| APPENDIX A..... | 123 |
| Consent Documents | |
| APPENDIX B..... | 127 |
| Testing Protocol | |
| APPENDIX C..... | 132 |
| Summary of Lesson Plans and Physical Aspects of the Room | |
| APPENDIX D..... | 138 |
| Instructions for Judges and Rating Form | |
| APPENDIX E..... | 142 |
| Descriptions of Judges' High and Low Ratings | |
| APPENDIX F..... | 147 |
| Individual Subject's Ratings | |
| REFERENCES..... | 155 |

LIST OF TABLES

| Table | Page |
|----------|------|
| Table 1 | 64 |
| Table 2 | 83 |
| Table 3 | 83 |
| Table 4 | 86 |
| Table 5 | 95 |
| Table 6 | 97 |
| Table 7 | 98 |
| Table 8 | 99 |
| Table 9 | 100 |
| Table C1 | 133 |
| Table C2 | 137 |
| Table F1 | 148 |
| Table F2 | 149 |
| Table F3 | 150 |
| Table F4 | 151 |
| Table F5 | 152 |
| Table F6 | 153 |

LIST OF ILLUSTRATIONS

| Figure | Page |
|--|------|
| Figure 1. Histogram of mean ratings of 1 st grade students' pretest improvisations..... | 90 |
| Figure 2. Histogram of mean ratings of 1 st grade students' post test improvisations..... | 90 |
| Figure 3. Histogram of mean ratings of 2 nd grade students' pretest improvisations..... | 91 |
| Figure 4. Histogram of mean ratings of 2 nd grade students' post test improvisations..... | 91 |
| Figure 5. Histogram of mean ratings of 3 rd grade students' pretest improvisations..... | 92 |
| Figure 6. Histogram of mean ratings of 3 rd grade students' post test improvisations..... | 92 |
| Figure 7. Histogram of mean ratings of total group's pretest improvisations..... | 93 |
| Figure 8. Histogram of mean ratings of total group's post test improvisations..... | 93 |

CHAPTER I

RATIONALE, PURPOSE, AND RESEARCH QUESTIONS

Rationale

As adult musicians, we play music. We work hard at it. We practice it. We perfect it. For a young child, the concept of playing music is just another variation on the age-old childhood occupation of play. Children play music for the joy of it—because it is there.

The educational community at large has studied the play of young children, and has developed theories of play: what it is, why it happens, what the child learns through play, and the role of play in the cognitive, social, and linguistic development of children. Music educators also have begun to look at the role of play in the development of musicality in children. This study sought to uncover the role of play as an instructional tool in the development of children's improvisational ability.

The Role of Play in Development

Freud (1942), Erikson (1950), and Csikszentmihalyi (1979) believed play to be the motivating factor in achievement. Freud believed that play might help a child overcome an unpleasant situation or the disappointment of unfulfilled wishes by being able to take control of life within the play paradigm. He believed sublimation of unfulfilled wishes and of unpleasant experiences is the motivating energy behind

the play of children. Erikson (1950) developed the idea further; believing play also propelled the child forward to new stages of mastering the environment. Erikson acknowledged the importance and uniqueness of play by maintaining that the child is able to master more of his or her environment in the unthreatening atmosphere created during play. Csikszentmihalyi (1979) suggested that play is visible when a child becomes totally absorbed in an activity, experiencing what he called “flow.” According to Csikszentmihalyi, flow is a positive psychological state that produces continued behavior in the absence of other reinforcers, and is a motivating factor throughout the life span. Flow is the goal of play, but only successful play results in flow.

Piaget, Vygotsky, and Bruner addressed the role of play in the development of children. Piaget (1962) proposed that children’s play occupied a central position in the development of intelligence. He theorized that the dual processes of assimilation and accommodation build intelligence. Assimilation is the process of adapting experiences and learnings to fit mental schemas already in place—to make new learnings fit in with old ones. Accommodation is the process of adapting existing mental schemas to fit with new experiences. He believed that play was the functional equivalent of assimilation, and imitation was the functional equivalent of accommodation, therefore giving play a pivotal role in the development of intelligence.

Piaget also theorized overlapping stages of play that corresponded, roughly, with his proposed stages of development: 1) practice play, corresponding with the sensorimotor stage; 2) symbolic play, corresponding with the pre-operational stage and containing the process of assimilation; and 3) play with rules—corresponding with the

concrete operational stage. According to Piaget, infants engage in practice play for the joy of mastery and to show off to themselves. Symbolic play, or imaginary play, begins at about age three and is dominant during the pre-school and early elementary school years. The last stage, play with rules, begins at about age seven and predominates play behavior until the age of eleven.

Other theorists and researchers have embraced Piaget's general theory of play, but have disputed the disappearance of play as the child matures. Lieberman (1977) and Sutton-Smith (1979) questioned Piaget's assertion that symbolic play is replaced by realism—or games with rules—in middle childhood. A study by van der Kooij and Meyjes (1986) on children aged four to nine found all types of play were evident throughout all ages studied. After a systematic review of the research on play, Smith, Takhvar, Gore, and Vollstedt (1985) agreed that the theory that symbolic play decreases in middle childhood is not supported by research findings.

Vygotsky, also disputing Piaget's ideas, contended that rules are inherent in symbolic play, and imagination is inherent in "games with rules." He believed that every imaginary situation contains rules in a concealed form—just as every game with rules contains an imaginary situation in a concealed form (Vygotsky, 1978). Central to Vygotsky's theory is a concept he calls the "zone of proximal development" (ZPD). He explained ZPD as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (p. 86). Vygotsky asserted that play is invented by the child at the point that

unrealizable tendencies appear in the child's development. Play expands the ZPD to allow the child to progress to more mature levels of knowledge and understanding (Vygotsky, 1966). He placed play on an equal footing with instruction, stating:

The play-development relationship can be compared to the instruction-development relationship, but play provides a background for changes in needs and in consciousness of a much wider nature. Play is the source of development and creates the zone of proximal development (Vygotsky, 1966, p. 16).

He further asserted that imaginative play does not die away at school age, but is able to permeate the child's attitude towards reality. Imaginative play is not as outwardly visible in the school age child, but is still very much present—now internalized as part of the thinking process. Vygotsky (1978) also believed play to persist throughout the life span, revealing itself in different forms at different junctures of life.

Jerome Bruner also contributed to the theory of play as it applies to cognitive development (Bruner, 1972; Bruner, Jolly & Sylva, 1976). Bruner recognized the developmental importance of practice play by which the child is able to learn the components of a skill. The components of a skill are gradually mastered through practice play, allowing for minimal attention to them when the skill is performed. Larger and larger components of a skill are capable of being mastered through practice play, which allows the child to respond to problem solving using these skills in a more efficient way (Bruner, 1972). Practice play, as opposed to just practice, is important because children can explore, discover and practice new behaviors while playing in an environment that is non-threatening—i.e. means-oriented instead of ends-oriented. Beyond practice play,

play in general allows the child to expand beyond the constraints imposed by the demands of reality and create new problem solving strategies and explore novel combinations (Bruner, Jolly, & Sylva, 1976).

The developmental play theories asserting that educational benefits are inherent in play behaviors have provided the impetus for researchers to determine if play does indeed have an effect on the development of children. Research began on play without a consensus on the definition of play, with each research project developing a definition of play for its own use (Smith, Takhvar, Gore & Vollstedt, 1985). In 1983, Rubin, Fein, and Vandenberg reviewed the existing research on play and developed a working definition of play comprised of these descriptors:

1. Play is intrinsically motivated.
2. Play is relatively free of externally imposed rules.
3. Play is carried out as if the activity were real.
4. Play focuses on the process rather than any product.
5. Play is dominated by the players.
6. Play requires the active involvement of the player.

A comprehensive search of play research by Fisher (1992) further delineated play into play that is controlled by the child (child-oriented play) and play that is controlled by an adult (adult-directed play). He also defined free play as being child-oriented play that takes place in a playroom equipped with toys under minimal adult supervision.

Fisher (1992) performed a systematic meta-analysis on 46 research studies written between 1974 and 1992 on the effect of play on the cognitive, linguistic, and social

development of children. Half of the studies were focused on the effects of play on cognitive development. The other half focused on the effects of play on linguistic or social development. Fisher found positive effects of play on all three areas of development—cognitive, linguistic, and social—with differing effects depending on the age of the child. The impact of play was shown to increase with age for cognitive and social skills, and was shown to decrease with age in the area of linguistic skills. Gains from adult-directed play were shown to be no greater than that for child-oriented play. Fisher also found that adult-directed play is fraught with methodological problems, specifically confounding variables introduced by varying adult behaviors.

Two longitudinal studies (Hartmann & Rollett, 1994; Sefer, 1995) showed the benefits of play on the development of children. Hartmann and Rollett (1994) implemented a study that infused play—both child-oriented and adult-directed—into the regular classroom instruction of 1st through 4th grade students. The treatment classes scored significantly higher on measures of intrinsic motivation, attitude, and divergent thinking than did the control classes not receiving the play treatment. Sefer (1995) looked at the influence of adult-directed play on students' divergent problem solving ability, motivation to learn, and expressive behaviors. The study consisted of a six-week experimental program for children aged 7 to 11, implemented each year for a 5-year period. The experimental group scored significantly higher on divergent problem solving, motivation to learn, and emotional expression than did the control group.

Play experiences in the classroom have been shown to positively impact children's cognitive, linguistic and social development. Is it possible that musical play experiences could also enhance musical improvisation in children?

The Role of Play in Musical Development

A few music education theorists have addressed the relationship between play and music composition and improvisation, and have suggested paradigms of music teaching appropriate to their theories (Addison, 1991; Ensley, 1987; Swanwick, 1988). Addison (1991) observed his own students involved in a play-oriented music curriculum, and then delineated characteristics common to both play and music. He also contrasted adults' and children's play, and suggested a method of instruction encompassing both adult-directed and child-oriented play.

Ensley (1987) based her writings on Huizinga (1950) and applied Huizinga's writings to the field of music. Huizinga believed play to be an essential part of the creation of culture and the development of civilization. Ensley correlated music to Huizinga's descriptors of play, then explored the value of play in musical education. Ensley proposed the playful creation of musical products should be at the center of music education.

An extensive treatise on the relationship between music and play was offered by Swanwick (1988) who made a case for music generation being a representation of play, basing his theory on Piaget's developmental theory. Swanwick proposed a model of musical development as well as a paradigm of instruction based on his theory.

Other music educators, citing general educational theory or anecdotal evidence, have called for the inclusion of play in music instruction during the elementary years (Addison, 1975; Andress, 1986; Campbell, 1989; Campbell & Scott-Kassner, 1995; Glover, 1990; Glover & Ward, 1993; Upitis, 1990, 1992).

Along with theoretical discourse on play and music, studies describing and analyzing the presence of play in musical improvisation have also been undertaken. Studies of playground song and ritual by Kartomi (1991) and Marsh (1995) found children participating in child-oriented play displayed improvisation behaviors—both individually and communally. These improvisatory behaviors were found to be comparable to the composition and improvisation practices of adults. Other researchers have documented episodes of playground singing, chanting, moving and the ritual involved with the performance and the peer-teaching of it (Merrill-Mirsky, 1988; Riddell, 1990). Riddell, in comparing children’s play music and classroom music, proposed that children develop in two parallel musical streams—one inside the classroom and the other within a play paradigm. She called for teachers to utilize music learning paradigms practiced by children during child-oriented play to enhance learning in the classroom. Also exploring the dualistic world of preschool children’s music, Veldhuis (1992) found differences in pitch, rhythm, and language usage between singing during play and singing elicited by adults. This seemed to amplify Riddell’s notion of two streams of parallel musical development and suggested that one stream was affected by adult expectations while the other included more spontaneous improvisation.

In their quest for understanding the role of play in musical development, musical researchers have also observed the musical play behaviors of preschool children. The Pillsbury Study (Moorhead & Pond, 1978) focused on the play behaviors of children, not only on the playground, but also within a classroom environment. The material, gathered over a period of 11 years, described the musical experiences—including improvisation—in which pre-school children participate if allowed to play in a well-equipped and supportive atmosphere. Follow-up studies by Shelley (1981) and Miller (1983) also documented the observed musical play behaviors of pre-school age children in a classroom setting. Although the extensive improvisation behaviors exhibited in the Pillsbury Study were not evident in these studies, both researchers inferred the positive value of participating in musical play experiences.

Purpose and Research Questions

Play has been given a prominent place in the educational theories of Bruner, Csikszentmihalyi, Erikson, Freud, Piaget, and Vygotsky. Although the theorists held differing views as to the role of play at specific ages, all agreed on the importance of play in the educational development of children. Educational researchers have documented the positive effect of play experiences on children's cognitive, linguistic and social development. This effect is shown to interact with the age of the child.

Musical play has been shown to produce improvisatory and compositional behaviors distinct from those observed in the teacher directed music experience. Some music educators have proposed embracing play as a part of music instruction to enhance improvisation and composition. However, it has not been established whether including

play experiences as a part of music instruction actually has an effect on the quality of children's improvisations.

The purpose of this study was to examine the effect of play experiences on the improvisations of 1st, 2nd, and 3rd grade children. Specifically, I sought to answer the following questions:

1. Is there a significant difference in the quality of musical improvisation of 1st, 2nd, and 3rd grade children receiving free play experiences during music instruction and the quality of musical improvisation of 1st, 2nd, and 3rd grade children not receiving free play experiences during music instruction?
2. Is there an interaction between grade level and the effects of free play on the quality of musical improvisation of 1st, 2nd, and 3rd grade children?

Definition of Terms

Play

For the purposes of this study, play will be free play—that is child-directed activities with minimal adult supervision in a classroom that is equipped with musical instruments.

Improvisation

For the purposes of this study, improvisation will encompass the musical performances given by children when asked to make up music.

CHAPTER II

REVIEW OF LITERATURE

Introduction

The research questions guiding this study involved play and its effect on musical improvisation. To provide a background understanding of play and its effect on musical improvisation, this chapter reviews selected works on play theory, play and cognitive development, play and music, and issues of measurement.

Play Theory

Early Play Theory

One of the first extended theories of play was the surplus energy theory. It was first proposed by Schiller in 1800 (Groos, 1898), then solidified by Spencer in his book Principles of Psychology (1896). The surplus energy theory proposed play to be merely the expenditure of energy the child has in excess of that needed to fulfill work and survival obligations. Although this theory had its merits, it did not take into account the child who plays beyond exhaustion and the very tired child who has a burst of energy when a play situation is suggested. Spencer also believed play, as well as being a release of exuberant energy, was the origin of all art—a theme that returned in other play theories (Erikson, 1950; Freud, 1942; and Huizinga, 1950).

Another complementary theory was the recreation theory put forth by Lazarus in 1883 (Mitchell & Mason, 1934). The recreation theory was based on the principle that a certain amount of rest and sleep are necessary, but beyond that, play was a means of recharging energy. To be effective in recharging energy, the play activity only needed to be a change from normal work activities. This was opposite of the surplus energy theory which looked upon play as a means of dispelling excess energy instead of providing a store of energy in and of itself.

Another play theory—the recapitulation theory—was proposed by G. Stanley Hall (Hall, 1906). Hall was heavily influenced by the theories of Darwin, and indicated children are the link in the evolutionary chain between animal and man. He believed children pass through the history of the race from protozoa to modern man. He proposed that the genetic typing of humanity is contained in each child, and that the childhood play years are merely a reenactment of life of the prehistoric and primitive man. The recapitulation theory helped further interest in children’s play behaviors and provided a theoretical base for the Orff methodology of music instruction for children (Warner, 1991).

Karl Groos, a professor of philosophy, studied the play of animals to better understand the play of man. He published two books: The Play of Animals (Groos, 1898) and The Play of Man (Groos, 1901) at the turn of the century. Groos concluded animals play because play practiced and perfected the survival skills they need as they mature. He theorized children play and imitate adults for the same reasons. Within the play of a specific child, he said, one could find the specific skills needed by that child

when he or she became an adult. Groos' work provided impetus for comparative research between the play of animals and of man. It also provided support for the growth of the belief that play served a function in the education of children.

Recent Play Theory

The psychoanalytic theory of play, supported by Freud (1942) and Erikson (1950), proposed that play was the outward manifestation of inward unrest. Freud believed play helped a child overcome an unpleasant situation or the disappointment of unfulfilled wishes by being able to take control of life within the play paradigm. He believed sublimation of unfulfilled wishes and of unpleasant experiences, which motivated humanity in general to seek knowledge, create art, and develop civilization, was also the motivating energy behind the play of children. Erikson developed the idea further, believing play propelled the child forward to new stages in the mastery of his or her environment. Erikson also delineated stages of young children's play: autocosmic, microspheric, and macrospheric. Erikson said autocosmic play began at birth, was centered on the infant's body, and was comprised, mostly, of repetitive actions and vocalizations. Microspheric play happened during the early toddler years. The child became absorbed in a contained world of toys and objects. The macrospheric stage occurred during the pre-school years. The child learned to play socially and also learned the appropriate times to play by himself. Each previous stage was believed to be incorporated into the subsequent stage—i.e. macrospheric play contained elements of both autocosmic and microspheric play. Erikson acknowledged the importance and

uniqueness of play—believing it was the immature form of the species’ ability to achieve mastery of its environment.

The cognitive-developmental theory of play was given its start by Piaget, and was at the center of his theory of intelligence (Piaget, 1962). Piaget theorized the dual processes of assimilation and accommodation built intelligence. Assimilation can be defined as the process of adapting experiences and learnings to fit mental schemas already in place—to make new learnings “fit” in with old ones. Accommodation can be defined as the process of adapting existing mental schemas to “fit” with new experiences. He believed play was the functional equivalent of assimilation, and imitation was the functional equivalent of accommodation.

Piaget also theorized overlapping stages of play that corresponded, roughly, with his proposed stages of development: 1) practice play, which corresponded with the sensori-motor stage; 2) symbolic play, which corresponded with the pre-operational stage and contained the process of assimilation; and 3) play with rules, which corresponded with the concrete operational stage. Infants are believed to indulge in practice play for the joy of mastery and to show off to themselves. Symbolic play, or imaginary play, is proposed by Piaget to begin at about age three and to be dominant during the pre-school and early elementary school years. Piaget also believed that the last stage, play with rules, began at about age seven and predominated play behavior until the age of eleven.

Other theorists and researchers have disputed Piaget’s theory of play. Lieberman (1977) and Sutton-Smith (1971) questioned Piaget’s assertion that symbolic play is replaced by realism in middle childhood. Van der Kooij and Meyjes (1986) studied 408

children aged four to nine, and found repetition, symbolic play and construction play remained high throughout all ages studied. Smith, Takhvar, and Gore (1985), after reviewing the literature, also agreed that the theory that symbolic play decreases in middle childhood was not supported by the research.

Vygotsky (1978) also disputed Piaget and contended that rules are inherent in symbolic play, and imagination is inherent in “games with rules.” He believed every imaginary situation contains rules in a concealed form—just as every game with rules contains an imaginary situation in a concealed form. He proposed a zone of proximal development (ZPD), which was defined as “the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers” (p. 86). Vygotsky asserted play is invented by the child at the point unrealizable tendencies appear in the child’s development and that play expands the ZPD to allow the child to progress to more mature levels of knowledge and understanding (Vygotsky, 1966). He placed play on equal footing with instruction, stating: “The play-development relationship can be compared to the instruction-development relationship, but play provides a background for changes in needs and in consciousness of a much wider nature. Play is the source of development and creates the zone of proximal development” (Vygotsky, 1966, p. 16). He further asserted that imaginative play does not die away at school age, but is able to permeate the child’s attitude towards reality. Imaginative play is not as visible outwardly in the school age child, but is still very much present—becoming internalized as part of the creative

thinking process. Vygotsky (1978) also believed play is a part of life throughout the life span, revealing itself in different forms at different junctures of life.

The presence of play beyond the childhood years was supported by Csikszentmihalyi (1979) who believed play is visible in the child or adult when he or she becomes totally absorbed in an activity, experiencing what Csikszentmihalyi called “flow.” Flow was defined as a positive psychological state that produces continued behavior in the absence of other reinforcers. According to Csikszentmihalyi, flow was a motivating factor in life. He asserted that flow is the goal of play, but only *successful* play results in flow.

Jerome Bruner has also contributed to the theory of educational play. Bruner recognized the developmental importance of practice play by which the child is able to learn the components of a skill. He believed that the components of a skill are gradually mastered through practice play, allowing minimal attention to them when the skill is performed. Larger and larger components of a skill are capable of being mastered through practice play, which allows the child to respond to problem solving using these skills in a more efficient way (Bruner, 1972). Bruner believed practice *play*, as opposed to just practice, is important because children can explore, discover and practice new behaviors while playing in an environment that is non-threatening—i.e. means oriented instead of ends oriented. He believed play in general allowed the child to expand beyond the constraints imposed by the demands of “reality” and create new problem solving strategies and explore novel combinations (Bruner, Jolly, & Sylva, 1976).

Defining Play

Research on play began without a consensus on the definition of play (Smith, Takhvar & Gore, 1985). The theories that propelled educational research on play dealt with the philosophical understanding of the motivations behind play and the purpose of play in the life of humankind. However, these theories did not provide a consensus of definition of play for the purpose of research measurement.

Smilansky (1968) proposed four classifications of play that were helpful in classifying what children do when they play: functional play, constructive play, dramatic play, and games with rules. Also, Parten (1932) developed a classification system for the describing the social aspects of what children do when they play: unoccupied behavior, onlooker behavior, solitary play, parallel play, associative play and cooperative play. Unoccupied behavior described the child who was totally uninvolved in the play experience. Onlooker behavior described the child who watched others, but did not overtly participate. Solitary play described the child who played independently without reference to other children. Parallel play described a child who played alone, but with activities or toys like those around him or her. Associative play described children playing together, but with no organization. Cooperative play described children who were organized for a specific purpose. In describing children's play, the categories of Smilansky and Parten have been used to provide classification of play during research situations (Rubin & Krasnor, 1980). Although these classifications have proven helpful in the research situation, they did not provide a distinction between activities that are play and activities that are non-play.

In 1983 Rubin, Fein, and Vandenberg did a comprehensive review of educational play literature and developed the following list as descriptors that distinguish play from non-play:

1. Play is intrinsically motivated.
2. Play is relatively free of externally imposed rules.
3. Play is carried out as if the activity were real.
4. Play focuses on the process rather than any product.
5. Play is dominated by the players.
6. Play requires the active involvement of the player.

This list was not meant to provide one definition of play, rather it was meant to provide overlapping criteria—the more of these indicators that were present, the more certain an observer could be in regarding the behavior as play.

Play and Cognitive Development

The theories of Freud, Erikson, Piaget, Vygotsky, Csikszentmihalyi, and Bruner that asserted that educational benefits were inherent in play gave impetus to researchers to conduct studies to determine if play does indeed have an effect on the educational development of children.

Several studies have addressed the effect of play on divergent problem solving in children. Lieberman (1965) explored the relationship between playfulness traits and divergent thinking—that is the ability to generate multiple solutions for a given problem. Lieberman, following the same rationale as Vygotsky (1966), maintained that play is not merely behavior, but is more of a predisposition that she called playfulness. She proposed

playfulness to be composed of five traits: 1) physical spontaneity, 2) social spontaneity, 3) cognitive spontaneity, 4) manifest joy, and 5) sense of humor. Subjects were 93 kindergarten children and were assessed by two raters on a Likert scale on the quantity and quality of the designated five playfulness traits. Also, for some reason that remains unclear, the children were rated with two questions on the raters' perception of the subjects' intelligence and physical attractiveness. Interrater reliability ranged from .66 to .83. The children were then given three tasks—two of which were adaptations of the Guilford and Torrance tests of creative thinking, and another divergent thinking task called the Monroe Language Classification Test. Ideational fluency scores were obtained from all three tasks. Two tasks yielded spontaneous flexibility scores, and another task yielded two originality scores: a weighted score and a cleverness score. Reliability estimates for the different measures ranged from .56 to .87. The author found significant correlations between all five playfulness traits and the three divergent thinking tasks, except for a correlation between physical spontaneity and originality. These findings provided impetus for further research on the link between play and children's development.

Dansky and Silverman (1973) chose to look at play's effect on associative fluency—a component of divergent thinking. The subjects were 90 preschool children of mixed gender and race. Random assignments were made to each of three conditions: play, imitation and control. Each child met with the experimenter individually in two meetings. The first was a five-minute rapport-building time with the experimenter. The second meeting lasted 10 minutes, took place a week later, and consisted of the treatment.

The subjects assigned to the play treatment were allowed to play freely with stimulus materials: paper towels, a screwdriver, a wooden board with 5 screws in it, paper clips, 3x5 cards, empty kitchen sized matchboxes, and a tray containing wet plastic cups. The subjects assigned to the imitation treatment watched an experimenter manipulate the same materials in various ways, and were then asked to imitate the experimenter. The control subjects were given four sketches and a box of crayons, and permitted to color as they wished. Immediately following the treatment, each subject was asked to name as many uses as possible for each of four of the experimental objects: the paper towel, screwdriver, paper clip, and match box. Answers were scored as to whether they were standard or nonstandard responses by two independent judges. Interrater agreement was 99% on the 1,549 answers. An analysis of variance showed the three treatments did not differ significantly on the number of standard responses for the four objects. However, an analysis of variance on the number of nonstandard responses showed the play condition having significantly more responses than the imitation or control conditions. There was no significant difference between imitation or control treatments. The authors stated: “The finding that exposure to the objects in the imitative context did not increase associative fluency above the base-line level of the control subjects appears to indicate that even active involvement with various stimuli does not insure an enrichment of associations to those stimuli” (p. 43). The authors, although demonstrating that free play facilitates associative fluency, were cautious about claiming any enduring effects of play on children’s abilities.

A study by Li (1978) followed the study of Dansky and Silverman (1973), integrating it with the playful mindset theory of Lieberman (1965). The design was similar to that of the Dansky and Silverman study, except, along with the free play, imitation, and control treatment conditions, another treatment condition was added—a make-believe condition. 120 children were randomly assigned to one of four treatment groups: make-believe, free play, imitation, and control. During the make-believe treatment, the children were told a five-sentence make-believe story. They were then presented with the stimulus materials and told “Let’s make believe or imagine that an object could become anything you would like it to be. Play with all of these things. Do whatever you would like to do with them” (p. 33). The children were then allowed a free-play time with the objects. The other treatment conditions followed the design of the Dansky and Silverman study. Immediately after the treatment, the children were asked to give as many uses for three of the objects to which they had previously been exposed, and for one object to which they had not been exposed. Standard and non-standard responses were measured separately. No significant difference was found between treatment groups for the standard responses for all four objects. However, significant differences were found in favor of both of the play groups on non-standard responses for one of the familiar objects, and in favor of the make-believe group on non-standard responses for the novel object. Non-standard responses for two of the familiar objects yielded non-significant results. The authors suggested the “introduction of make-believe elements in the instructions seems to have greater facilitating effect than just asking

subjects to play with the objects. One might infer a possible transfer effect of a make-believe or imaginative playfulness” (p. 35).

In 1975, Dansky and Silverman expanded their study. This time, they changed the control group to an “intellectual” group, giving clues about the physical properties and uses of the objects, requiring the subject to guess which object the experimenter was talking about. They also used a different set of objects for the associative fluency test than had been used for the treatment conditions. Subjects were 36 preschool children, randomly assigned to each of the three treatments: free play, imitation, or intellectual. In this study, the play treatment produced significantly more responses for both standard and non-standard uses than did either the imitation or intellectual groups. There was no significant difference between the imitation and intellectual groups. These findings supported their previous work and also suggested that “playful activity can provide children with an opportunity to organize their experiences and exercise their cognitive abilities in a manner that is likely to facilitate imaginative adaptations to future situations” (p.104).

Pepler and Ross (1981) investigated the role of convergent and divergent play on divergent thinking with a series of two studies. The first study used 64 three- and four-year-old children, equally divided by sex and age. The children were randomly assigned one of four treatment conditions: (a) play with convergent materials, (b) play with divergent materials, (c) observe convergent activity, and (d) observe divergent activity. The children were seen individually for three 10-minute sessions and then given a battery of problem-solving tasks approximately four hours after the last play session. Their

individual play sessions were analyzed from videotape for six categories of behavior. Interrater reliability estimates ranged from .70 to .87. The play behaviors of the children in the convergent task group tended to focus on the implied task inherent in the materials. The children in the divergent play group displayed a wider variety of behaviors: investigation, grouping, and symbolic play. The evaluation tasks consisted of two divergent and three convergent problem-solving tasks. Interobserver reliability estimates ranged from 83.3 to 100. The data from the divergent evaluation tasks revealed the children with divergent play and observing experiences scored higher on fluency—with the divergent play group scoring higher also on the originality measures. For the convergent evaluation tasks, the data showed no differences between play and observe or convergent and divergent treatments.

Pepler and Ross questioned the validity of the convergent evaluation measures, as they did not closely match the convergent play experience. Another study was undertaken. This study was a replication of the first study, with some changes. The changes implemented were:

1. The group of students who just observed was replaced with a non-play control group.
2. The order of two divergent and four convergent tasks were counterbalanced and administered by an experimenter who was unaware of the treatment group in which the child belonged.
3. The convergent tasks were redesigned to more closely utilize skills inherent in the convergent play materials.

An analysis of the play behaviors supported the findings of the original study. Also, the divergent play group gave more unique responses than either the convergent play group or the control group. The analysis of data for the convergent evaluation tasks revealed the convergent play group used more strategies in their problem solving. The convergent group also tended to persist with strategies that appeared reasonable, but would not work. The divergent group tended to use more trial and error type strategies, and would not tend to stay with a strategy that was not working.

In conclusion, the researchers stated: “The comparison of the effects of play with convergent and divergent materials suggested that the effects of convergent play experiences were very specific, whereas the divergent play experiences transferred more generally” (p. 1210).

Vandenberg (1981) was interested in the impact of play on problem solving as it related to the child’s age. In this study, subjects were 90 children—30 aged 4-5, 30 aged 6-7, and 30 aged 8-10. To control for perceptual motor abilities, the children were given the Block Design Subtest from either the WPPSI or WISC, depending on their age. The children were ranked and, in a randomized block procedure, assigned to either the experimental or control group. The children were seen individually for the study. Both experimental and control groups were allowed a 10 minute period of play with Legos. The play group was then allowed to play with the experimental materials—eight notched sticks and six pipe cleaners—for 10 minutes, and the control group was shown the experimental materials and asked questions about them for 10 minutes. The play group was scored for the number of constructions, the number of purposeful manual acts, and

how often the child performed acts specific to the experimental tasks. These scores were combined to provide a “richness of play” score for each child in the play condition. After the play or question session, the children were given two tasks, which required them to use the experimental materials as tools. The tasks were scored based on the number and level of hints the child needed to complete the task, and the number of goal oriented acts performed. The researchers found the play to be most beneficial for the 6 to 7 year-old group. They did not find any correlation between richness of play and task performance. The authors concluded play is most beneficial for performance of tasks commensurate with the child’s abilities. They hypothesized play, rather than providing specific skill practice, provides a general schema, or a background of experience, that is used in a specific task.

A series of studies looked at the effect play has on convergent problem solving. Convergent problem solving involves solving a problem where there is one correct solution. Smith and several colleagues did these studies involving a lure-retrieval paradigm (Simon & Smith, 1985; Smith & Dutton, 1979; Smith & Simon, 1983; Smith, Simon, & Emberton, 1985). The lure-retrieval paradigm generally required subjects to receive some treatment—play, training in problem solving, or no treatment—then figure out how to retrieve a valued object by constructing a tool out of available materials.

Smith and Dutton (1979) used 108 four-year-olds randomly assigned to three groups: play, training, and a control group. All subjects met with experimenters individually. The play condition subjects were shown the experimental materials: three 9-inch, three 6-inch and three 3-inch sticks, and five blocks. The children were shown that

the sticks were of different lengths and fit into holes in the blocks which could link them together. They were then allowed to play freely with the objects for a total of 10 minutes. The children then immediately attempted two problem-solving tasks.

The training group also received information about the objects and had a two-minute period for exploratory manipulation of the sticks and the blocks. The training subjects then watched the experimenter join the sticks together, and repeated the experimenter's actions. The training subjects were also asked to join the sticks together in order of length and given help by the experimenter by such questions as: "Which is the longest?" "Which is the next longest?". The training procedure lasted for a total of 10 minutes including the introduction, exploratory time, and training period. The child then immediately attempted the two problem-solving tasks.

The control group was divided into two sections of 18 children each. Control group one attempted task 1 and task 2, control group two attempted task 2 only. Task 1 required the children to link two of the longest sticks together to retrieve a lure—a procedure that was a part of the training treatment. Task 2 required innovative problem solving—linking three of the longest sticks together to retrieve the lure. This procedure was not addressed in the training treatment. Standardized hints were given if the child ignored the problem for 1 minute. The fifth and final hint virtually solved the problem for the child. If a child was able to perform the task without hints, it was called a spontaneous solution.

Groups were compared with an analysis of variance for solution time, number of hints needed, and the number of spontaneous solutions, motivated solvers, and

unmotivated solvers. The data showed no significant difference on task 1 between play and training groups, and a significant difference between the treatment groups and the control group with the treatment groups scoring better than the control groups. However, on task 2, which required more innovative problem solving, the play group required significantly less time and significantly fewer hints than the other two groups. Also, the play group contained significantly more spontaneous solvers, significantly more motivated solvers, and significantly fewer unmotivated solvers. This data suggested play is more beneficial than training or no play on innovative problem solving and enhanced motivation for problem solving.

Cheyne (1982) criticized this research design. Cheyne held the previous experiment was tainted by researcher bias. Smith and Simon (1983) replicated the 1979 Smith and Dutton study, but incorporated stricter controls over experimenter bias. This revised study supported the findings that play and training experiences were equivalent, but failed to support the theory that play is superior to training in more innovative tasks.

In 1985, Smith, Simon, and Emberton attempted to replicate this series of studies involving lure-retrieval. This study also supported the findings that play and training experiences were equivalent, but failed to support the theory that play is superior for more innovative tasks. The authors concluded their report by questioning the efficacy of “using a single short experimental session to demonstrate learning or innovative effects of play experience” (p. 106). This questioning led to a more complete review of the research paradigm involved in testing the effect of play on convergent problem solving.

In 1985, Simon and Smith looked systematically at the design of research studies investigating play and problem solving. In a review of literature, they found 11 published studies that related to play experience. After analysis of all 11 studies, they pronounced the general research designs flawed. In 5 studies, scoring bias was possible; in 9 studies, testing bias was possible; and in 10 studies, treatment bias was possible. The study they subsequently undertook attempted to control for these flaws. A more exact procedural design was implemented, researcher bias was controlled, and four groups were studied: play, training, passive experience of materials, and alternative materials (drawing). The data showed no significant differences between any of the groups. The lack of significant difference for even the training group prompted the researchers to question this research paradigm as a valid measurement of the effect of play or training on problem solving. They explored several possibilities for the failure of the research. The most plausible reason, in their eyes, was the possibility that 10 minutes, once, is not enough time for “play” to really happen and they suggested longer periods of time to fully appreciate the benefits of play.

A study by Curran (1993) provided more support for the use of longer play periods in researching the role of play in cognitive development. Curran stated young children use mental representations in play they are not yet able to verbalize. She hypothesized children would score better on cognitive tasks if the tasks were approached within a pretend protocol—also that children that had a lot of experience with pretend play would benefit more from the opportunity to deal with problems under a pretend condition. Subjects were 123 children who were rated by their teachers as to frequency of

pretend play, and then randomly assigned to two conditions: a pretend condition where tasks were presented within a pretend state, and a control condition where tasks were presented without reference to pretend states. The assessment tasks were a pretend-reality task, an appearance-reality task, a fluency task, tasks involving conservation of number, liquid and volume, and two subtests of the Stanford-Binet: bead memory and copying tasks. The pretend condition produced a significant overall superior effect. When tasks were looked at individually, the pretend condition did not, however, assist children in solving conservation tasks. Frequency of pretend play, as indicated by teacher rating, was an indicator of superior performance on all tasks, whether or not the subject was a part of the pretend condition or of the control group. This research supported the use of regular, longitudinal play experiences over isolated play experiences.

As discussed in the rationale, Fisher undertook a systematic meta-analysis of play research in 1992. A thorough search yielded 1,171 studies on some aspect of play conducted since 1974. Fisher grouped these into these subgroups:

1. Play as dependent variable—how other factors influenced play (796 studies).
2. Psychological play therapy (295 studies).
3. The effect of play on the development of children (81 studies).

Of the 81 studies on the effect of play on the development of children, 46 were included in Fisher's statistical meta-analysis. Nineteen of the articles not used were discarded because they were purely theoretical or discursive reviews, lacking any usable statistic. Sixteen of the articles not used were discarded because they were untranslated foreign language articles, unobtainable dissertations, or "simply impractical to locate by

any reasonable means” (p. 162). Half of the 46 studies were focused on the effects of play on cognitive development, which Fisher delineated as creativity and problem solving. The other half focused on the effect of play on linguistic or social development.

Fisher’s coding procedures in the area of methodological features included (a) sample size, (b) subject characteristics, (c) treatment conditions, especially how subjects were selected and assigned, and (d) research design and validity weighting. The coding procedures in the area of substantive features included (a) type of play condition, (b) stimulus materials used, (c) setting, (d) duration of treatment, (e) measurement technology, (f) dependent variable, (g) statistical test employed with degrees of freedom, (h) alpha levels, and (i) an effect size outcome.

Through systematic analysis, Fisher found support for the claims of the individual studies that play positively influences cognitive, linguistic and social development in children. The magnitude of effect obtained was $r=.347$, which falls between the categories of modest and note-worthy in Cohen’s general guidelines for statistic use in the behavioral sciences (Cohen, 1977).

Although Fisher’s meta-analysis (1992) of the literature declares play to positively affect cognitive development, he states the lack of consistency and agreed upon research paradigms in the experimental studies have made the effects less than robust. In another less formal review of these types of studies, Smith (1995) states:

...the experimental situations employed were just not ecologically very valid ways of trying to demonstrate any functions of play. The experiments were usually short—about ten or fifteen minutes—whereas any benefits of play might

be slow and cumulative; the experiments involved “asking” or “requiring” children to play, when supposedly we were interested in benefits of free play. The experiments were clearly susceptible to experimenter effects, so any apparent findings were not robust (p. 14).

Many research findings on the effects of play have been tentative because of the brevity of the play treatments. There have been, however, a few studies done that were able to look at the long-term benefits of play. A longitudinal study was conducted in Vienna from 1983-1987 (Hartmann & Rollett, 1994) to examine the long-term effects of a program of study entitled the “Viennese Play Curriculum.” Twelve experimental classes were matched by age, gender, IQ, and socioeconomic status with twelve control classes. The control classes received “regular” instruction. The experimental classes received the Viennese Play Curriculum during the students’ 1st through 4th grade educational process. Experimental teachers were trained in the pedagogy of play, experimental classrooms were equipped with appropriate materials and surroundings, and teachers were counseled to schedule approximately four hours per week for free play as well as incorporating non-specific amounts of didactic, teacher-guided play into the curriculum. After the 4 year project, children were tested with divergent thinking measures, as well as measures for emotional attitude towards school, social behavior, achievement motivation, and scholastic achievement. Teachers were also measured for their emotional attitude towards school. The experimental group, as hypothesized, scored significantly higher than the control group on measures of divergent thinking at the end of treatment. They also scored significantly higher on most measures of

emotional attitude towards school, social behavior, and achievement motivation.

Teachers in the experimental group also scored significantly higher in measures of emotional attitude towards school. There was no difference between the experimental groups and the control groups on scholastic achievement. This finding was positive, because some had feared losing at least four hours per week of instruction time over a period of four years might negatively impact achievement. This fear was not supported.

The authors highly recommend inclusion of play in the daily curriculum stating:

...all of these benefits for the play-enriched classes were obtained without any forfeits in scholastic achievement. Students who could devote themselves to play activities showed a more positive attitude toward learning and a greater intrinsic motivation than the controls. Children of the play group found more divergent ways to solve problems and produced more creative ideas (Hartmann & Rollett, 1994, p.202).

Another longitudinal study of play oriented curriculum was implemented in Yugoslavia from 1985-1990 (Sefer, 1995). It looked at the influence of play on students' ability to produce creative products, motivation to learn, and expressive behaviors. It consisted of a six-week experimental program administered to 99 children aged 7 to 11, implemented each year for a 5 year period. The program was revised annually through consultation with implementing teachers. The program used an interdisciplinary approach—combining science, math, drama, history, music and art—and included games, improvisations, small group play, individual free play, and special research. Although the teacher structured much of the activity a priori, there was allowance for student choice of

method and content, and teachers strove to remain open to student initiated method. The children were randomly divided into a control and experimental group. The control group received “regular” instruction, and the experimental group participated in the experimental 6-week program each year. Assessments were made several times each year. For assessment, teachers were interviewed based on a standardized set of questions about implementation and effectiveness of the program, and the children were given multiple problem solving tasks including individual, small group and large group tasks. The tasks ranged from visual art to writing assignments to group drama. The task evaluations were done qualitatively and quantitatively on a series of previously defined protocols reflecting the assessment media evaluated. The evaluators were various professionals outside the research system:

Outside observers and evaluators included two professional writers, two professional artists, five experts for taped group activities, seven experts for taped class activities and trained university students for in-class and in-group observations. Outside evaluators did not know if the class observed was experimental or control (Sefer, 1995, p. 9).

The data gleaned from the experiments was analyzed through statistical techniques and content analysis was used for the essay materials gathered throughout the study. The experimental class showed significant superiority in art production for the first three grades, and in story writing in the third and fourth grades. The other measures showed constant significant superiority of the experimental group in the variables of creativity, fluency, and originality. On non-parametric measures, all observers agreed the

process of work in the experimental classes was superior to that of the control classes. The experimental classes were superior in developing fluent, original, complex ideas; product effectiveness; and emotional expression. The authors highly supported this type of instruction in the elementary grades, and have been tracking both the control and experimental students' progress in subsequent years. Preliminary data from this post experimental tracking pointed to the long-term effects of this curriculum being even more dramatic than the short-term effects.

Play and Music

The studies of play and its effect on musical development reviewed in this section can be divided into two subsections: those studies focusing mainly on the theoretical understanding of the role of play in musical development, and those studies seeking to provide quantification of the role of play in musical development.

Play and Music: Theoretical Considerations

A few music theorists have explored the similarities between play and music and proposed paradigms of music teaching appropriate to their theories. This section discusses musical play theories as developed by Addison (1991), Ensley (1987), and Swanwick (1988).

Addison (1991), after experiencing the benefits of a play oriented curriculum in his own teaching, and assuming the reader's acknowledgment of the benefits of play, delineated the characteristics of play and music and found them to be similar. The characteristics he proposed as being in common are:

1. Both are pleasurable.

2. Both are self-contained—they are of value in and of themselves.
3. Both are voluntary activities.
4. Both invite total absorption and a suspension of reality during performance.
5. Both are symbolic.
6. Both require immense effort and a “seriousness” of purpose.
7. Both are voluntary—the beginning, progress and end are in control of the players.

In order to understand children’s unique perspective on play, Addison contrasted children’s and adults’ play, coming up with four contrasts he believed were important:

1. Children’s play is active, while adults’ play tends to be passive.
2. Adults are more competitive, while children will often play just for the joy of playing.
3. Child’s play operates with fluidity, moving from make-believe to construction of objects, to rough and tumble with ease. Adult’s play tends to be fixed in routine.
4. For children, play is life. It is a very important part of daily routine and necessary for healthy development. It is, for many adults, a decoration on life—pleasurable, but not indispensable.

Addison used these differences to make suggestions for utilizing the benefits of play during music instruction. He suggested:

1. The activity level—movement, dancing—must remain high.
2. The means should be valued over the end.

3. The structure should be guided by the child—moving fluidly between activities without pressure for completion.
4. The exploring and handling of musical objects is of utmost importance.

Addison's work was highly anecdotal and minimally grounded in research, but had the advantage of being based on extensive personal experience in play oriented methodology. Other published articles by Addison added descriptive data to support his theories (Addison 1972, 1975, 1985).

At a deeper theoretical level, Ensley (1987) reviewed Johan Huizinga's book, Homo Ludens: A Study of the Play Element in Culture and applied Huizinga's writings to the field of music. Huizinga, a historian, established a commonality between law, war, poetry, and art based on their inclusion in what he called a "play-sphere" (Huizinga, 1950). Huizinga believed play was the essence of civilization, and defined the nature and function of play. Ensley (1987) correlated music to these descriptors of play, stating the following elemental attributes were common to both music and play:

1. Both were outside the reasonableness of practical life.
2. Both were intrinsically motivated.
3. Both were determined by values that transcend logical ideas, visibility, and tangibility.
4. Both required a special set of designations to express their values.
5. Both functioned in society as contest and representation.
6. Both were defined and limited by the cultural context in which they occur.
7. Both had roots in sacred ritual.

After correlating the descriptors of both play and music, Ensley then explored the value of play in education. Her reasoning followed that of Bruner, seeing play as developing the ability to transfer learning—providing the player with the flexibility to use behavior experimented with in play in new situations.

Ensley made a case for the presence of play motivations and behaviors during the musical behaviors of composition, performance and listening. She also suggested that the times of growth in the development of music in the Western world were driven by novel ideas introduced by performers and composers embodying a playful paradigm. Ensley submitted the presence of play in music and the importance of play in civilization and education and called for a shift in philosophy to stress playfulness in the musical education process.

Ensley drew these conclusions:

Recognition of a play element in music, rather than merely emphasizing creativity, makes creativity a required objective of music instruction. Play, necessary for the expression and production of culture, is essentially creative in nature. The playful and imaginative manipulation of musical ideas will yield creativity, which is the mode of culture production. Thus music instruction should provide opportunities for students to assume active roles in the playful manipulation of musical ideas in order that creativity might emerge...

Recognition of the functions of play in music can serve to elucidate the arrangement of learning sequence, the evaluation of musical learning, the securing

of a high level of motivation, and the creation of a favorable social climate in the classroom. (p. 43)

Perhaps the most complete treatise on the relationship between music and play can be found in a book by Keith Swanwick: Music, Mind and Education, published in 1988. In his book, Swanwick discussed the role of the arts in society. The paradigm he embraced is that of the arts as play. He drew on the theories of Piaget and offered three play impulses: a) mastery, b) imitation, and c) imaginative play. The mastery impulse is analogous to practice play—repetition for the joy of accomplishment. Swanwick defined the other two impulses thusly: “In imaginative play, the world around us is transformed to our standards; in imitation, we are transformed, we become like something else” (p. 42). The play impulses outlined by Swanwick were found, by him, to be impulses driving music also. The mastery impulse can be found in the development of technical skills—physical technique, notational reading, and sound manipulation. This mastery impulse can also be found in the appreciation of the virtuosity of others. The imitation impulse can be seen, musically, as the recreating of music from score, in performing or composing programmatic pieces, or in developing the emotional impact of the piece. The impulse of imaginative play can be seen in music as the creation of structure transformations—the “novel re-constitution of musical possibilities” (p. 57).

Swanwick elaborated further, citing Piaget’s theory of intelligence that placed a balance of imitation and imagination at the source of intellectual growth. Swanwick postulated the same balance of imitation and imagination—plus mastery—is at the source of musical growth. Swanwick also proposed another level, which he calls meta-cognition.

He proposed that if meta-cognition occurs, it only occurs in musicians beyond the age of 15 and consists of “self-awareness of the processes of thought and feeling in a value-response to music” (p. 74). This level of development has the capability of subsuming all previous levels of development.

To further support his theory, Swanwick gathered 745 compositions of 48 children aged 3 to 11, over a period of 4 years. Swanwick defined composition thusly:

Composition takes place when there is some freedom to choose the ordering of music, with or without notational or other forms of detailed performance instruction. Others may prefer at times to use the terms improvisation, invention or ‘creative music’. All of these fall within this very broad definition of ‘composition’, the act of assembling music (p. 60).

From an analysis of these compositions, Swanwick proposed eight developmental modes that act and interact with each other in a spiral fashion. The eight modes he proposed were the sensory, manipulative, personal, vernacular, speculative, idiomatic, symbolic and systematic modes. These modes are used in combination with the concepts of mastery, imitation, imaginative play, and meta-cognition to produce the musical developmental whole.

The sensory mode is the mode primarily used by children up to age 3. They are concerned with the sensuousness of sound, particularly timbre. It is characterized by the disorganization of musical elements and experimentation with sound sources.

The manipulative mode begins appearing about age 4 and 5. Regular pulse is organized and interest in technique arises. Technique is explored in relationship to the

physical structure of the instrument (i.e. glissandi on keyboard instruments), and the music tends to be repetitive.

The personal expressiveness mode becomes visible in compositions of children aged 4 to 6. Changes in loudness and tempo are explored, and there are hints of musical phrases beginning to develop. There is little feel for structure, however, and the music sounds as if it is being performed intuitively without pre-planning.

The vernacular mode begins to appear around age 5, and is more solidly established by age 7 or 8. It is characterized by shorter compositions that contain elements of structural organization. Phrases are evident, as are meter, ostinati, and syncopation. Compositions in this mode tend to be predictable and imitative.

Although sometimes found earlier, the speculative mode is most apparent in children between the ages of 9 and 11. Novelty is valued, surprises are planned, and children seem absorbed in making their piece “just right.” Experimentation with structural ideas is also apparent.

The idiomatic mode is most often recognized in children of 13-14. The child seems to try to emulate accepted popular performers and their pieces often strongly resemble popular music. Musical authenticity to style is important. The speculative mode from which the child has just come becomes overwhelmed by the need to imitate.

The symbolic mode is rarely seen before age 15, and in some people is not seen at all. The music takes on a strong personal significance and students operating in this mode have a high commitment to music that appeals to personal feelings. The music takes on a function of symbolism for the self and the self's value system.

The systematic mode is only experienced by the sophisticated, highly developed music maker. He or she is able to think about music in a highly organized way. Music may be approached from a historical, musicological, psychological or philosophical paradigm. Compositions coming from this mode may be influenced by novel approaches, or research into various functions of music or compositional styles. Swanwick also stated: "In the systematic mode, the universe of musical discourse is expanded, reflected upon, discussed and celebrated with others" (p. 80).

Although speculative in presentation, the theory developed by Swanwick offered an alternative mode of thinking about the musical development of children and adults based on play theory.

Quantification of the Relationship Between Play and Music

Merrill-Mirsky (1988) looked at elementary school children's musical play behavior over a period of 5 years in Los Angeles. She documented 342 variations of 117 handclapping and ring games, jumprope chants, and cheers. The music was discussed as oral tradition, and was analyzed with reference to the development of gender roles, and musical variation between and within ethnic groups. Although the study did not focus on the educational benefits of this musical tradition, it did provide a documentation of the rich and varied music outside of the regular school music tradition.

Riddell (1990) also did a descriptive study of children in Los Angeles. Like Merrill-Mirsky, she was interested in documenting the music and movement of the children, but Riddell was also interested in the learning and teaching process involved in transmission of this tradition. She studied 16 singing games of fifth-grade children, and

then interviewed the children in an open interview format. She also supplemented this data with interview of two other children and two adults that had grown up in the neighborhood and were familiar with the tradition. She found even though music on the playground is complex with highly idiomatic performance requirements, children were able to function as teachers and learners of this music without adult intervention. Riddell found several discrepancies between this playground teaching and learning and classroom music education. She found children were actively involved in playground music in the role of critic even when they are not actively performing. She also stated children do not break the whole into components when teaching and learning, preferring to start over at the beginning if a mistake is made. Riddell proposed that music teaching and learning outside the classroom operates in a holistic fashion, while that of the music classroom is approached from a reductionistic standpoint. Riddell called for music educators to incorporate more of this holistic natural learning mode for children into their methodology.

Kartomi (1991) gave a descriptive analysis of several case studies she undertook in Australia between 1969 and 1980. She compared her compilation of children's song to the Western concept of improvisation and composition. She concluded that much music making is done within the playground setting and is analogous to the adult concept of improvisation. This improvisatory musical play process was undertaken both individually and collectively on the playground. The style of the improvisations resembled that of the structured songs the children learn from adults or each other, but the improvisations were quite often performed as if "thinking aloud." Melody, tempo, dynamics, meter, and

rhythm were all open to variation within the improvisations. The author stressed the importance of children's improvised song as a valid musical creation that expressed the experience and purposes of the creating child, and encouraged the education community to embrace this improvisatory mode in the classroom setting.

Marsh (1995) also did a descriptive study of children's playground singing games to gather information on improvisation and composition and children's meta-cognition of these processes. The subjects were multiethnic children between the ages of 5 and 12. They provided her with 448 examples of playground singing games and concurrent interviews. She found children were very capable of extensive improvisation and composition within this tradition. The compositional processes were generally collaborative between members of the play groups. Marsh suggested this arrangement facilitates the ability of children of varying levels of musical maturity to work together—allowing the less mature to have their knowledge and skills expanded. The interviews also suggested children have an understanding of their own compositional processes, and are able to discern form and melodic/rhythmic problems and work on solving them within a musical situation. Marsh called for more collaborative work in the classroom, as well as for giving children the opportunity to improvise upon classroom musical material.

Collectively, these studies support the theory of the existence of a musically valid and separate stream of music that children are capable of experiencing without formal adult instruction. Through experimentation and peer instruction, children were able to learn music in a play setting.

Veldhuis (1992) was interested in the differences between spontaneous play song and adult initiated song. In order to compare these two types of singing he studied the spontaneous singing of 29 four-year-old children and compared language maturity and pitch range to that found during elicited singing in the presence of an adult researcher. It was found that the language in the spontaneous singing was less mature than that of the elicited singing, and the spontaneous singing was performed at a significantly higher pitch than that of the elicited singing. Veldhuis offered support for the theory in which two separate streams of development in childhood are hypothesized—that of playful spontaneous music making and that geared towards adult expectations. This study implied the natural stream of spontaneous improvisation of young children could be accessed to assist in their musical education experience.

The spontaneous improvisation of young children and the musical growth naturally evolving as a result of it was the focus of the Pillsbury Study, which was undertaken over a period of 11 years—from 1937 to 1948 (Moorhead & Pond, 1978). A school was established in 1937 for children ranging in age from 18 months to 8½ years, with most of the children aged 3 to 6. Enrollment was generally between 10 and 20 students. The students were given free access to a wide range of musical instruments, recordings, art materials, books, pictures, and play props. Formal music periods were not held, and although the students were supervised by teachers, there was a minimum amount of structure and children were left to their own devices for musical or non-musical activities, either indoors or out. Teachers recorded observations of the children on 5 x 8 index cards each day. Although the teachers sought to nurture the children's

exploration and improvisation , they tried to not impose their ideas and plans upon the children (Pond, 1981). The teachers participated with the children in the following modes:

1. Participation in rhythmic dialogue—teacher played a pattern and the student answered back, or the student played a pattern and the teacher echoed or provided a variation.
2. Talking with the child about what he or she was doing, making the child consciously aware of what was being done spontaneously.
3. Playing and improvising with the child whenever asked.
4. Providing the children with technical skills and information whenever help was required.
5. By observing children's discussions and joining in when the teacher's presence was requested.

The findings were published in four volumes intermittently from 1941 to 1951: Volume I, Chant; Volume II, General Observations; Volume III, Musical Notation; and Volume IV, Free Use of Instruments for Musical Growth. In 1978, all four volumes were combined and reprinted as Music of Young Children (Wilson, 1981). Most of the reported work centered on improvisation and composition—both vocally and instrumentally—and gave insight through anecdotal reports into the compositional practices of this stage of life. From becoming a part of the children's everyday musical life and from analyzing the data collected, the researchers concluded young children have an innate sense of musicality, are capable of musical growth and development through

free play activities, and exhibit compositional and improvisational behaviors worth nurturing (Moorehead & Pond, 1978).

Also in the area of young children's innate musicality, Shelley (1981) reported on a research project that took place at the University of Maryland's Center for Young Children during a 10 week period in 1978. The author stated:

“The goals of the study were:

1. to show that natural musicality was expressed in the Center for Young Children's setting;
2. to confirm observations made in the Pillsbury Foundation School, 1937-1948;
3. to develop and refine observational techniques; and
4. to collect and analyze raw data about spontaneous musical behavior observed in young children” (Shelley, 1981, p. 28).

Subjects were two intact classes—one containing 13 three- and four-year-olds, and the other containing 17 four- and five-year-olds. A musical specialist instructed the children during 30-minute sessions once a week. During this group music time, the children were introduced to percussive and melodic musical instruments as well as improvisational musical experiences with the voice and body. Observations were made twice weekly during the regular daily one-hour free play time. An observational tool was developed specifically for this project. The following behaviors were observed 20 times or more between the two classes, or at least 10 times in one of the classes:

1. Goes to music center by choice
2. Observes but does not participate

3. Examines and manipulates instruments
4. Selects record and plays it on the phonograph
5. Accompanies record with instruments
6. Moves with accompaniment
7. Sings without accompaniment
8. Combines music and visual aids.

Of perhaps equal interest are the behaviors that were not observed frequently:

1. Is guided to music center by teacher
2. Sings, or chants with accompaniment
3. Chants or moves without accompaniment
4. Exhibits social frustration
5. Exhibits musical frustration
6. Invites others to join
7. Demonstrates skill repetition
8. Evaluates own performance
9. Evaluates and refines performance.

The researchers found children do explore music in an environment conducive to experimentation, and the data also gave support for these findings of the Pillsbury Foundation Study in relation to the characteristics of children's spontaneous music making:

1. It is an embryonic type of polyphony and is linear in structure.

2. It resembles music of primitive peoples in choice of interval and freedom of rhythm.
3. It is expressed through chant. (Shelley, 1981, p. 52)

Although some of the findings of the Pillsbury Foundation Study were supported, it is clear that much of the social and performance behaviors that were a significant part of the Pillsbury School did not take place during this study. It is possible the ecological circumstances, or individual cultural differences of the children were responsible for this discrepancy.

Shelley also cited problems in funding technological aspects, found much of the recording of the children's music making chaotic, and suggested in-depth analysis of technological method to enhance further studies.

Miller (1983) implemented a study with a naturalistic design that consisted of observation of 95 children aged 3-, 4-, and 5-years old during two 30-minute periods. Children were observed in a large area with musical materials readily available. They were able to play freely, either alone or with others, or to ask the teacher to play any record they desired. During the first observation time, behaviors were categorized according to previous researchers' lists of behaviors. The observed behaviors were analyzed and formatted into a document called the *Musical Behavior Observation Matrix* which included all previously observed behaviors and provided means to correlate the behaviors to age, sex, and race. The second observation was videotaped and analyzed by two observers. Inter-observer reliability was .92. From this study, several conclusions were drawn:

1. Young children are capable of making music freely and spontaneously in their natural settings by:
 - a. participating consistently within a designated time period.
 - b. examining and manipulating instruments.
 - c. requesting records to be played.
 - d. identifying preferences for listening.
 - e. listening attentively.
 - f. accompanying recordings while playing instruments or singing.
 - g. singing, chanting, or moving body parts with recorded music.
 - h. creating songs with random pitch sequences, creating extemporaneous pitches on melodic instruments or creating rhythms on non-melodic instruments or creating rhythms on non-melodic instruments.
 - i. imitating spontaneous rhythms and movements of peers.
 - j. responding to the basic beat of recorded music.
 - k. demonstrating increased responses to music with fast tempi.
 - l. experimenting with combinations of sound.
2. Young children are capable of responding musically without teacher intervention.
3. Young children possess social behaviors that lend themselves to interacting appropriately.

4. Imitation increased social interaction in the form of talking, laughing, and additional imitation.
5. Three-year-olds imitate their peers less than 4 and 5 year olds.
6. Three-year-olds engage in solitary and parallel play more than 4 and 5 year olds. (Miller, 1983, p. 110-111)

These findings support behaviors observed by Shelley (1981) and Moorehead and Pond (1978). However, just as in the study by Shelley, extensive compositional behaviors as reported in the Pillsbury Foundation Study were not reported. Again, it is possible these behaviors were not observed because of ecological considerations or individual differences in the children.

Littleton (1991) did an observational study of preschool children, analyzing their cognitive and social play behaviors in a house play setting, and their vocal, movement, and instrumental play behaviors in a free play music setting. A videotape of the children at play was made and data collected at one-minute intervals. The researcher also provided anecdotal descriptions of some of the play sessions for amplification of the quantitative data. More functional and constructive play was exhibited in the music setting, and more dramatic play was exhibited in the house setting. The music setting also promoted twice as much solitary play as did the house setting. The data was further analyzed according to gender. In the music area, both boys and girls showed preference for instrumental play. When the children demonstrated music behaviors in the house area, girls were found to participate in movement more often, boys were more often found to be involved in vocal and instrumental behaviors.

The studies cited above show evidence of a separate stream of music in child's play that contains a rich and varied repertoire, performance practice and peer transmission of that repertoire and performance practice. The observation of pre-school children's play practices in the area of music also support the ability of children to be involved in music making at their own initiative. In the Pillsbury Study, that natural music making was shown to evolve into more sophisticated forms when allowed to take place over a longer period of time. Follow-up studies of pre-school children's free play musical behavior has supported the Pillsbury Study's findings that children are able to develop musical activities in which to participate if left to their own devices. The studies have not, however, substantiated the findings of the Pillsbury Study in relation to social and improvisational behavior.

Issues of Measurement

Amabile (1983) proposed a method called the consensual assessment technique by which to measure the quality of creative products. She argued that expert judges from a particular domain, when providing subjective judgment, impart construct validity to the judgment. She further asserted that the supposedly objective nature of criterion based assessment is not truly objective, as the test designer relies on subjective opinion as to what constitutes the ultimate criteria. Amabile believed an honestly subjective instrument is a more valid indicator of quality of a creative product than a subjective measure veiled in objectivity.

Consensual Assessment

Amabile tested her theory of assessment on a series of 21 visual art and creative writing tasks. She used a variety of different types of judges, ranging from practitioners, to teachers, to laymen. She also used a variety of subjects from early childhood to adult. Interjudge reliability for the different tasks ranged from .72 to .93, satisfactorily reliable in Amabile's judgment.

Amabile stated that for proper utilization of the consensual assessment technique, the assessment task must result in a clearly observable product, it must be open-ended enough to permit flexibility and novelty, and the task should not depend on special technical skills that would be perhaps more highly developed in some subjects than in others. Requirements delineated for the judging process were:

1. Judges should all have experience within the domain in question. The experience does not have to be the same, but they should have had enough experience to have developed implicit criteria for creativity within the domain.
2. Judges should judge independently of one another.
3. Judges should not be trained or given a set of criteria by which to judge.
4. To keep creative quality separate from other aspects of the pieces, judges should be required to make judgments on other aspects besides creativity, and instructed to keep these dimensions separate in their mind.
5. The products should be rated relative to each other, and not against a set standard within the domain.

6. Each judge should view the products in a different random order, and assess each dimension required in a different random order. (Amabile, 1983, pp. 38-39)

Consensual Assessment Applied To Musical Assessment

This consensual assessment paradigm was applied successfully to musical assessment by Bangs (1992), Brinkman, (1994), Hickey (1995), and Webster and Hickey (1995). Bangs (1992) adapted Amabile's (1983) consensual assessment paradigm for use in evaluating 3rd grade students' compositions for a pre and post test. He developed 19 five-point Likert items for use in the judges' rating modeled after Amabile's Dimension of Judgments for Artist-Judges (1982). Nine items related to the creativity of the compositions, and 10 items related to other musical aspects of the composition. Each composition was given a mean creativity rating derived from the judge's ratings. The interjudge reliability among the three judges used was .76 for the pre-test and .82 for the post test, indicating a reliable test.

Brinkman (1994) also used the consensual assessment paradigm for rating two compositions from each of 32 high school students. Using a 7-point Likert scale, each of three judges was asked to rate each melody on originality, craftsmanship, and aesthetic value. Judges were not given explicit definitions of these terms, and were expected to use their own implicit understanding of these three descriptors. Interjudge reliability ranged from .77 to .96, indicating a reliable test.

Hickey (1995) used a consensual assessment rating tool to rate compositions of 21 fourth and fifth grade children for purpose of comparing high creativity process and low

creativity process. A 7-point Likert scale was used on three items asking for global ratings of creativity, craftsmanship, and aesthetic appeal. Estimates of interjudge reliability between the three judges on the upper third and lower third of the group were acceptable, ranging from .62 to .93. If all twenty-one subjects were included in the reliability estimates, however, the estimates dropped to between .18 and .80. Hickey reasoned the greatest lack of agreement was between judges in the mid-section of the grouping. Because of this, Hickey used the upper 1/3 and the bottom 1/3 of the subjects' scores for the purpose of comparing creative process differences. Hickey found the use of the consensual assessment technique to be a reliable method for rating children's compositions. In analyzing the three items on the rating instrument, all three dimensions correlated highly with each other. The correlations were: aesthetic appeal and creativity, .87; aesthetic appeal and craftsmanship, .92; and creativity and craftsmanship, .81.

Webster and Hickey (1995) were interested in the appropriateness of different types of rating scales for the assessment of children's compositions. The researchers, in reviewing the literature that assessed children's musical compositions, found eight rating scales that assessed the quality of compositions. In presentation of the rating scales used, two approaches to style and two approaches to content were found. The two approaches to style were termed explicit and implicit. Explicit style contained extensive descriptions of what was to be rated. Implicit style was vague to allow the evaluator to decide on criteria. The two approaches to content were specific and global. The specific rating scales contained items on musical characteristics such as rhythm or harmony; while the global rating scales contained items that rated such broader concepts such as originality

or appeal. Webster and Hickey developed two parallel forms incorporating differing combinations of implicit/explicit style with specific/global content. These rating scales were used by four expert judges to rate one set of ten children's compositions. Statistical analysis of the judges' scoring revealed both forms were reliable. However, the implicit/global rating scale was the most reliable in assessing the more global dimensions of creativity and aesthetic value. Webster and Hickey (1995) stated "It might be reasoned by some researchers that explicitly designed rating scales have greater reliability because the judge is given a clear sense of the item's meaning. Results of this study using four judges' ratings of ten children's compositions suggest otherwise." (p. 36).

CHAPTER III

METHODOLOGY

Introduction

Research methods involving human subjects and real life are always less than ideal. The experiment is always a compromise between stringent control and the reality of the diversities of life (Myers, 1972). The goal of measuring children's behavior in as natural a setting as possible must be balanced with the goal of controlling for as many undesirable confounding variables as possible. The design of this study was developed to control for as many of these variables as possible.

Design of the Study

The design of the study was guided by the following research questions:

1. Is there a significant difference in the quality of musical improvisation of 1st, 2nd, and 3rd grade children receiving free play experiences during music instruction and the quality of musical improvisation of 1st, 2nd, and 3rd grade children not receiving free play experiences during music instruction?
2. Is there an interaction between grade level and the effects of free play on the quality of musical improvisation of 1st, 2nd, and 3rd grade children?

A pretest posttest control group design was chosen for this study. All students were tested at the beginning of the semester by being asked to make up music to go with the folk story “Goldilocks and the Three Bears.” Their improvisations were videotaped for later review. For the next 14 weeks of the semester, half of the students received some time during their regular music instruction for free play with the classroom music instruments in the music classroom, while the other half of the students did not receive free play time during their music instruction. At the end of the semester, all students were tested by again making up music to go with “Goldilocks and the Three Bears.” These improvisations were also videotaped. All taped improvisations were submitted to three judges for rating on the dimension of quality on a 5-point Likert scale. Judges were unaware of whether the improvisations were from the pretest or post test, in which grade the student was enrolled, or whether or not the student had participated in the free play experience. These ratings were then subjected to an analysis of variance ($\alpha=.05$) to see if the free play experiences had an effect on the measured quality of the improvisations.

Threats To Research Design

Campbell and Stanley (1963) cite history, maturation, testing, instrumentation, statistical regression, experimental mortality, and the reactive effects of experimental arrangements as possible threats to research design. Although they emphasized that “from the standpoint of the final interpretation of an experiment and the attempt to fit it into the developing science, every experiment is imperfect” (p. 34), they stressed the importance of recognizing possible threats to the integrity of research and addressing those insofar as

is possible. These possible threats were considered when developing this study's protocol.

History

The threat of history centers on the events occurring between the first measurement and the second measurement. In this case, the ratings of the musical improvisations of "Goldilocks and the Three Bears" at the beginning of the semester and the musical improvisations of "Goldilocks and the Three Bears" at the end of the semester constitute the first and second measurement. This threat was addressed by randomly assigning children to each group, therefore any events affecting specific children would be randomly dispersed among the groups.

Maturation

Maturation poses a threat in that the students' change in performance between the pre and post test might be influenced by merely the passage of time rather than the free play experiences. This threat was addressed by including a control group in the design.

Testing

The effects of taking a pretest upon the scores of a post test could prove to be a threat to the research, as could the possibility that taking a pretest might increase or decrease the responsiveness to the treatment—in this case, the free play experience. The first possible testing threat, that the experience of the pretest could possibly affect the scores on the post test was addressed by having both groups—free play and control—take both tests. The second possible testing threat—that the pretest could make the students more aware of improvisation during the intervening time between pre and post testing—

was a possibility that might need to be taken into account in interpreting data. To lessen the impact of the improvisation task, it was designed to be close to other classroom experiences with which the children were already familiar. Although the students had not made individual improvisations in music class during the previous semester, the music teacher had included large group and small group activities that required the students to make up music to go with folk stories. Also, to avoid having the students pre-plan their post test improvisations during free play or music class, students were not told they would be doing another individual improvisation task at the end of the semester.

Instrumentation

Instrumentation threat can occur when changes are made in the measuring instrument, or when changes occur in the judges. This possible threat was addressed in several ways. The same testing instrument was administered to all students for both the pre and post tests. The same five administrators gave both sets of tests. The administrators were trained as a group in general testing procedure. The rooms in which the students were tested were empty classrooms of similar shape and size in a wing that was awaiting renovation. All testing rooms were set up in the same configuration and had the same instruments available for use by the students. The students were randomly assigned to administrators. Students were brought out in random order from music class for testing. Those not currently testing carried on the same activity—group singing of songs already known by the children—to provide similarity of activity from which the student would be leaving.

Possible instrumentation threat from changes in judges was addressed in several ways. All judges rated all improvisations. Judges were not aware of which improvisations were from the pretest and which were from the post test. Judges were also unaware of which children received free play experiences, and which did not receive free play experiences, or the grade level of the child. Each judge received the improvisations compiled in a unique random order, and viewed and rated the improvisations on one day. Although judge fatigue might have affected the rating of the improvisations, the presentation of the improvisations in unique random order to each judge helped address this threat.

The improvisations were videotaped (which provided audio as well) as opposed to merely audiotaped. This procedure may have allowed the judges to be influenced by the appearance of the child during performance. However, use of the video was able to provide a more complete observation of the improvisation, a benefit that was believed to outweigh any possible biases.

Statistical Regression

The threat of statistical regression is possible when groups have been selected on the basis of their extreme scores. To address this possible threat, all students in the 1st, 2nd, and 3rd grades in the school were invited to be in the study. The school did not have entrance requirements. Of the total student body, 92% of the first graders, 100% of the second graders, and 100% of the third graders took part in the study.

Experimental mortality

Experimental mortality may threaten the study when there is loss of subjects from a group. This proved to be problematic for this study, as two 1st grade children moved away during the semester. These students' pretest ratings were dropped from the data set, and the appropriate statistical adjustments were made to accommodate unequal sample sizes.

Selection Bias

Selection bias may threaten the study when students perform differently because they feel special for having been chosen to be a part of one group or another. This threat was addressed in several ways. It was not possible for students to not know that their group was doing something different in music class than the other group. However, all classes were taught by the same person in the same classroom, and the same curriculum was followed except for the free play experiences received by some of the children. The children not receiving free play experiences received experiences that enriched the regular music curriculum during the corresponding time. The experiences were designed to be ones the students found enjoyable, but the experiences did not focus on composition or improvisation.

Reactive effects of experimental arrangements

Reactive effects of experimental arrangements may occur when the experimental situation is not one found in real life situations. This possible threat was addressed by keeping the music schedule as normal as possible, by using an improvisation task for testing purposes that was not abnormal to an elementary music classroom, by using the

regular music teacher to implement the free play experiences, and by not specially designing a supporting curriculum to facilitate the improvisatory experience.

Statistical Considerations—Sample Size

The research questions contained two areas of inquiry: differences between groups in the measured quality of children's improvisations due to experiences received during instruction, and the interaction of grade with any differences. The sample size in each group—play and non-play—began with 55 children. Because of experimental mortality, the play group finished the study with only 53 students, with both missing students from the 1st grade. According to Hinkle, Wiersma, and Jurs (1988), the unequal sample sizes could have an effect on the findings if the variances differ between groups. If the larger group has a larger variance, the test will be too conservative. If the larger group has a smaller variance, the test will be too liberal. If the variances between the two groups are the same, the differing sample sizes' effect is minimal. To understand the effects of differing sample size in this study, variances between groups were compared.

The sample size for this study was large enough to detect an effect size of .75 with the power of the test being .90 at an α of .05 for comparison between the total play and non-play groups. However, to compare the interaction of grade on any differences between groups, the sample size recommended in each of 6 groups (2 treatments X 3 grade levels) at the same level would be 52. This would result in a total sample size of 312 students. This study contained 108 students. The negative aspect of using a small sample size used will be considered when data is discussed.

Research Setting

Because the research questions addressed the use of free play as *part* of the instructional process, the ideal setting for the study was an elementary school that would allow free play to be integrated into its regular music curriculum with all the students in 1st, 2nd, and 3rd grades available for random selection. This implied a music curriculum already in place. Further, the school's music program would ideally be taught by a music specialist, be well-equipped with instruments appropriate for musical free play, have adequate space to accommodate free play, and have adequate instruction time in music to allow some time for free play activities. Several schools were approached for hosting the study. The school chosen to host the research met these requirements and was also able to accommodate block random sampling of students by grade level without disruption of the normal scheduling, which allowed a more natural setting in which to implement the treatment. The school chosen was a small town elementary school in the southwestern United States. Approval from the Institutional Review Board for Human Subjects Research was received as well as permission from school officials. Information letters, as well as forms requesting permission from parents were sent home with each child (see Appendix A for consent documents).

Subjects

The school had 37 students in the 1st grade, 36 students in the 2nd grade, and 40 students in the 3rd grade. All 1st, 2nd, and 3rd grade students were invited to be a part of the study. The consent percentage for each grade level was as follows: first grade-92%, second grade-100%, and third grade-100%. Students were assigned through block

randomization by grade level to either the control or treatment group. All students' names for grade level one were placed on separate pieces of paper in a container. Separate pieces of paper on which the numbers one through ninety-nine were written were placed in another container. A name was drawn from the first container, and a number was drawn from the second container and assigned to that name. This was done for all names in the first grade. Using the random number generator contained in the SPSS computer based statistical package (1998), the first 17 numbers corresponding to student assigned numbers were assigned to the class receiving free play experiences (half of the total group), and the rest were assigned to the class not receiving the free play experiences. The same procedure was then repeated for grades two and three, assigning half of the students to the class receiving free play experiences and half to the class not receiving the free play experiences. Group sizes were as follows:

Table 1

Group Sizes

| | Control | Free play |
|-----------------------|---------|-----------|
| 1 st grade | 17 | 15 |
| 2 nd grade | 18 | 18 |
| 3 rd grade | 20 | 20 |

The inequality between the group sizes of the 1st grade free play and control groups was caused by 2 children moving away from the school district during the semester in which the study took place. These students were dropped from the study.

Method

All students were tested with a researcher-designed instrument to elicit improvisations at the beginning of the study (see Appendix B). This test required the students to “make up” music to accompany the story of “Goldilocks and the Three Bears”. Music was requested of the student for four specific points during the story: when the bears go for a walk, when Goldilocks falls asleep, when Little Bear is very angry, and when Goldilocks runs away. To prepare for the improvisation task, the administrator of the task picked up the student from the music class and took him or her to a separate room containing classroom music instruments. The student was given instructions, then left alone to work on the music for as long as he or she wanted. When the student signaled he or she was ready, the administrator rejoined the student and read the story while the student provided improvisations at the appropriate time. Each reading/improvisation was videotaped for later rating by three independent judges.

Over the next 14 weeks, all students received regular music instruction. This instruction was an eclectic mix of singing, movement, instrument playing, music reading, listening, and performance preparation (for a more complete description of the regular music instruction and the physical aspects of the music room, see Appendix C). Fourteen weeks comprised most of the spring semester, allowing for pre and post testing and special days and events at the school. All music classes met during the mornings. During each two-week interval, students were scheduled for 120 minutes of music instruction. The students experiencing the free play sessions used 20 minutes of the 120 allotted minutes of instruction for free play, resulting in seven 20-minute free play sessions.

Those students not experiencing the free play sessions did enrichment activities during the corresponding time that followed the regular curriculum. These enrichment activities were designed to be enjoyable for the children, but did not focus on improvisation or composition.

The existing schedule of the music classes at the school helped maintain a somewhat normal music routine while allowing half the children to have free play experiences. In this school, general music class and library instruction were alternated while the homeroom teacher had a preparation period. For example, while one first grade class was in general music, the other first grade class was at the library. The following day, the groups would switch, allowing equal time for all students in both the library and general music class. To allow for block random sampling by grade level for purposes of this study, the students were grouped for library and music classes during this semester by free play and non-free play instead of by homeroom teacher.

At the end of the 14 weeks, all students were tested in the same manner as they were before treatment using the same researcher-designed instrument to elicit improvisations. This testing was also videotaped.

Development of the Improvisation Task

In order to measure the quality of a child's improvisation, a task needed to be designed that would provide a valid example of the child's work. According to Amabile (1983), to provide a valid example of an original product—in this case an improvisation—a task must result in a clearly observable product, must be open-ended enough to permit flexibility and novelty, and should not depend on special technical skills

that would be perhaps more highly developed in some students than in others. Amabile also stressed that the procedure to elicit an original product needs to avoid giving the student the expectation of evaluation or of being watched during the creative process. The choice in how to do the activity should also carry as few restrictions as possible (Amabile & Gitomer, 1984; Hennessey, 1989).

Improvisation Task

A protocol to elicit improvisation was developed according to these guidelines. The task resulted in a clearly observable product—a videotaped improvisation. Four points during the story were delineated as needing music, but the student was not told which instruments to use, or how the music should sound or how long it should be to allow for flexibility and novelty. All of the instruments available to the child for performance were easily playable, and not of the type on which some students would be taking instruction that would give them a technical advantage. The physical situation was designed so that the student could be alone while he or she worked on the improvisation. The students were not given time limits or expectations, they were merely told to let the testing administrator know when they were done. To avoid having the student feel as if he or she was being evaluated, the administrator joined in the performance of the improvisation by providing a dramatic reading. The children were not told that their improvisations would be rated. If the children asked why they were being videotaped, they were told that the administrator was collecting children's music in something like a scrapbook to see what children could do.

Field Test I

The first field test was given to 7 children not involved in the final study—three first graders, two second graders, and two third graders. Each child was scheduled to arrive separately at 1-hour intervals. The children were received into a room that contained all the instruments that were to be found during the free play experiences of the final study, and were allowed to explore the full range of classroom instruments as long as they desired before being tested. The exploration time ranged from 20 to 45 minutes. After exploration, each child was taken individually by the researcher to a separate room that contained each type of instrument that was to be available for use during the main study's improvisation task. The instruments required minimal skill level to manipulate musically and were of the normal classroom variety. The researcher asked the child if he or she knew the story of "Goldilocks and the Three Bears". All students were familiar with the story. The child was then asked to compose music to go with the story. Different points in the story of "Goldilocks and the Three Bears" were used to focus the compositions while still providing the freedom of an open-ended task. Four scenes in the story were delineated as those scenes needing music (see Appendix B for specifics on the music requested of the children). The child was then told that the researcher would leave so that the child could work on the music as long as he or she wanted to. After the child signaled the improvisation was ready, the researcher and the child participated together in a performance of "Goldilocks and the Three Bears" that was videotaped.

The field testing procedure followed the basic protocol developed, except when children seemed unclear on instructions, and then further instructions were given. These

modified instructions were noted for revision. The procedure was videotaped for later review. During the review, these problems were noted:

1. Some of the wording of instructions seemed confusing.
2. The placement of the video camera accommodated the visual requirements of the test, but did not fully facilitate the audio requirements.
3. Some students forgot for which part of the story they were supposed to compose music.
4. Students needed some sort of cues during performance to know when to play their compositions.

After revisions were made to the improvisation protocol, the improvisation task was then tested on five more occasions using a total of 51 first, second, and third grade children that would not be involved in the main study. The problems observed during the first field test were rectified in the following ways:

1. *Some of the wording of instructions was confusing.* Some examples of problems included the administrator's request to "compose music", which was changed to "make-up music". Also the children wanted to know which instruments they were to use, so the script was revised to specifically state that all instruments could be used. These revisions were made, along with others, which resulted in less confusion being expressed by the children.
2. *The placement of the video camera accommodated the visual requirements of the test, but did not facilitate the audio requirements.* The video camera was placed closer to the child. During the first series of improvisations, the

researcher believed the camera would be distracting if placed too close to the child, and so the camera was placed unobtrusively 10 feet away from the student. The camera did not appear to be distracting for any of the students at this distance, but the audio reception was inadequate. For the subsequent series of tests, the camera was placed on a tripod 6 feet from the student. The camera caught the profile of the student. This revised camera placement did not seem to be distracting to the student, and both the visual and audio recording was adequate for scoring purposes.

3. *Some students forgot for which part of the story they were supposed to compose music.* Some pictures were made for the scenes that needed music and left with the student during the composition process. After this adjustment was made, all children remembered the four scenes requiring music.
4. *Students needed some sort of cues during performance to know when to play their improvisations.* During performance, the appropriate pictures were pointed out to cue the students as to when to play their music. This seemed to alleviate the confusion as all students were able to play their music on cue.

The revised script for the improvisational task is contained in Appendix B.

Development of the Free Play Protocol

The free play experience was designed to fit within the definition of play as given by Rubin, Fein, and Vandenberg (1983). These guidelines are listed below, with their operational implications for this study delineated.

1. *Play is intrinsically motivated.* This was operationalized by providing the opportunity and encouragement to play, but not requiring individual children to play.
2. *Play is relatively free of externally imposed rules.* The children were given no instruction on the “correct” way to play, sing, move, or interact beyond that which was necessary for the safety of the children and the instruments, and the facilitation of social harmony.
3. *Play is carried out as if the activity were real.* This indicator of play is contained mainly within the mind of the player. Its presence or lack thereof is not discernible by the teacher and was not the focus of any intervention.
4. *Play focuses on the process rather than any product.* Students were not asked to produce anything during their free play time.
5. *Play is dominated by the players.* Students were allowed to control the content and process of play within safe and socially harmonic boundaries.
6. *Play requires the active involvement of the player.* Although participation was not required, students were encouraged to become involved in the play activities.

Field Test II—Free Play Protocol

To see if 1st, 2nd, and 3rd graders would participate in free play in a music classroom over an extended time, free play experiences were provided once a week for a semester (18 weeks) at a school similar to the one at which the main study would be. It was also a small town elementary school in the southwestern United States, and had a

total population of 186 children in the 1st, 2nd, and 3rd grades. The students were randomly divided in half, as they would be during the main study. However, because of scheduling difficulties, the students not participating in the free play session during Field Test II participated in organized folk dancing and group singing instead of regular music class. Also in Field Test II, the researcher administered the free play sessions instead of the regular music teacher. The students were allowed a full 45-minute class period for play, once a week, for a full semester. During the first play session, children were given guidelines for classroom procedures and then allowed to play for the rest of the period. Children were told that they had free access to any and all materials in the classroom with which to play as they wished, and that the teacher would be available for help, if desired. Five minutes before the end of the class period, students were directed in clean-up activities and organized for dismissal.

The free play sessions were videotaped and reviewed by an independent reviewer to verify that children had participated in free play over the semester. The reviewer verified that the six indicators of play given by Rubin, Fein, and Vandenberg (1983) were operationalized as planned. She also verified that children participated in free play activities and that for most students, interest in play seemed high for the entire semester. Activities in which the students engaged included instrument and sound exploration, figuring out songs by ear, singing, moving, and making up music alone and in groups. Some students spent time pretending to teach each other in a music class, taking turns being the teacher. Some students dressed up wearing as many instruments as they could and then danced to make music. The students also organized concert bands, marching

bands, and rock bands with the appropriate conductors, baton twirlers, or lead singers. Several children linked together the headsets and microphones for the 3 portable pianos available to them and pretended to be airplane pilots making a variety of sound effects with the pianos and their voices. A major preoccupation that happened every free play session was drumming together in groups ranging in size from two to eight.

During the first three free play sessions, four students showed distress over the high noise level in the classroom by putting their hands over their ears. Two other students complained to the researcher about the noise level. The children complaining were agreed with by the researcher that the noise level was quite high. They were told that lots of noise sometimes happens when children are playing. The students were then guided to an area of the room that was not quite as noisy, and told that if it was still too uncomfortable for them after a few minutes that we would try and figure out something to do about the noise. The students who had complained joined back in the play and did not express any more distress about the high noise level. After the third session, there were no students that indicated that they were uncomfortable with the noise level in the classroom.

Not all of the students' time was spent on playing with the instruments to "make music". Some children practiced cheerleading, others used jingle sticks as cars and had races. Intervention by the researcher was needed intermittently to halt activities that were possibly unsafe for the children or the instruments. Examples of these activities included mock "sword fights" with rhythm sticks, playing chase in the classroom, playing the maracas on other children's heads, standing on each other's backs and constructing

human pyramids, and experimenting with the sound produced by hitting the acoustic piano with various objects. Participation was very high, with only 9 children showing intermittent non-participation throughout the semester. These non-participation episodes all took place during the last half of the 45-minute period, and the children stated they were bored, or just didn't want to do anything. Non-participating children were approached by the researcher who attempted to direct them to other activities—sometimes successfully, sometimes unsuccessfully.

All classifications of play outlined by Parten (1932)—onlooker behavior, solitary play, parallel play, associative play, and cooperative play—were observed during each play period. There was also functional play in the exploration of sound and the waving around of sticks and other instruments, and construction play in the planning out of group compositions, disassembling and assembling of instruments, and the building of makeshift race tracks for the jingle stick cars. Dramatic play infused many of the play situations. Organized games with rules were not evident. Perhaps if musical “board games” were available for play, children would have been observed participating in these organized games with rules.

Main Study—Free Play Protocol

For the main study, the regular music teacher was given instruction on the operationalization of the 6 play indicators. She was instructed to not “teach” but merely to facilitate social harmony and safety of both the students and the instruments. The free play sessions were videotaped with an unmanned, unobtrusively placed video camera and reviewed by the researcher to ensure the music teacher implemented the free play

experience according to the guidelines delineated. All guidelines for free play were operationalized in the main study as delineated, and the play activities were similar to the activities engaged in by the students in Field Test II.

In the main study, the free play sessions took place during regularly scheduled music class periods in the regular music classroom. During the first play session, children were given guidelines for classroom procedures and then allowed to play for the rest of the period. Children were told that they had free access to any and all materials in the classroom with which to play as they wished, and that the teacher would be available for help, if desired (for a description of the classroom environment and materials available to the children, see Appendix C). A few minutes before the end of the play period children were organized for dismissal. Each subsequent treatment period consisted of 20 minutes of free play with a short time for organized dismissal at the end.

Refining the Testing Procedure

Before undertaking the testing of students in the main study, certain questions needed to be answered:

1. How long would testing of so many students take considering the travel time between the classroom and the testing sites?
2. Are there unforeseen problems with testing in a public school environment?
3. Are there any technical problems that might occur running so many different videocameras?
4. Would testing administrators not familiar with the improvisation task have trouble administering the test?

Field Test II—Testing Procedure

To be sure that the testing procedure would run smoothly in the main study, students that were taking part in the field test exploring the administration of free play were also given the improvisation task. The task was administered to all 1st, 2nd, and 3rd grade students—both the play and the non-play—at the end of the semester.

The administrators were assembled 1 hour prior to beginning of the testing. They were 6 female volunteers that ranged in age from 23 to 42. The administrators had little music experience, but all had experience working with young children. They were given packets that contained instructions, the script for “Goldilocks and the Three Bears”, pictures for cueing the children as to when to play their improvisations, and a checklist of students that had been pre-assigned to them from each class being tested. The assignment of student to administrator as well as the order in which the children would be tested had been predetermined through block randomization by class via a randomly generated number chart. The administrators were taken through the procedure, and given time to ask questions. They then watched the researcher administer the test with one of the administrators role-playing the part of the student. When all questions had been answered, the administrators went to their own testing area, which had been previously set up with the appropriate equipment. The researcher then checked with each administrator individually to make sure she knew how to operate the video camera and understood the procedure. The students were then tested over the next 5 days.

The testing of the students revealed that each administrator could be expected to test 2 to 3 students in a 45-minute class period. The public school environment did yield

some difficulties. Some of the videotape equipment owned by the school that was scheduled to be used was not actually available. Also one of the planned testing sites was not available for the entire time. Student absences also needed to be accounted for when scheduling testing. The video cameras did not pose any special problems, except for automatic shut-off mechanisms that needed to be monitored. The testing administrators demonstrated no problems, and seemed to be comfortable with the testing procedure.

Main Study—Testing Procedure

For the main study, a different set of administrators was used because the main study was located 280 miles from the Field Test II site and travel was not possible for the first set of administrators. The main study's administrators were 5 female volunteers aged 22-56. They, also, had little experience in music, but had extensive experience working with young children. Training was undertaken in the same manner in which it was done during Field Test II, with additional information included on the automatic shut-off mechanism that was a part of 2 of the video cameras. Other steps were also taken to alleviate some of the problems encountered in Field Test II. Videotape equipment was utilized that did not belong to the host school to ensure its availability. The testing sites were checked and double-checked to make sure they were not expected to be available for someone else during this time. To accommodate student absences, two days per class were allotted for testing. If a student was absent when he or she was scheduled, the originally planned administrator tested the student at a later date. As there were only absences that interfered with scheduling on the first testing day, those absences were made up during the next time the students were in music class. The pretest was

administered over a period of four consecutive days at the beginning of the semester and the post test was administered over a period of four consecutive days at the end of the semester.

Rating of Improvisations

Each student produced two separate video clips: an improvisation from the beginning of the semester and an improvisation from the end of the semester. Each student's improvisational rendition of the four scenes during the performance of "Goldilocks and the Three Bears" was considered for rating as one composition. The spoken performance by the test administrator was edited from all tapes to limit judge fatigue. Also, any other improvisations the child chose to make outside of the four target improvisations were edited out of the tapes.

Each video clip was assigned a number, and then placed in order onto videotape via the order of numbers found on a randomly generated table. Each of three judges received the videotaped improvisations in a unique randomized order. The improvisations were rated for quality on a 5-point Likert scale. The change in quality ratings between the beginning of the semester and the end of the semester for the two groups were then subjected to an analysis of variance to determine whether the free play experiences had a significant impact on the quality of the students' improvisations.

Consensual Assessment Technique

The rating system used in this study was modeled on the consensual assessment technique developed by Amabile (1983). In the consensual assessment method of rating, the rating of quality is performed by a group of independent expert judges who rate each

product via a Likert-scale on a specified dimension. The final rating of the product is the mean of the expert judges' individual ratings. Expert judges from a particular domain, when providing subjective judgment, impart construct validity to the judgment. The consensual assessment technique was developed specifically to measure creativity through the rating of creative products. This study, however, chose to rate the product without inferring the presence or lack thereof of an underlying construct called creativity. The guidelines, therefore, were altered slightly to accommodate the rating of the quality of the product instead of the creativity of the product.

The improvisation task was developed along the guidelines delineated for this type of assessment as discussed previously. For this type of assessment, criteria was also established for the judges and judging protocol. Requirements delineated for the judging process as given by Amabile (1983) were:

1. Judges should all have experience within the domain in question. The experience does not have to be the same, but they should have had enough experience to develop implicit criteria for the measured dimension within the domain.
2. Judges should judge independently of one another.
3. Judges should not be trained or given a set of criteria by which to judge.
4. Judges should be required to make judgments on other aspects besides the targeted dimension, and instructed to keep these dimensions separate in their mind.

5. The products should be rated relative to each other, and not against a set standard within the domain.
6. Each judge should view the products in a different random order, and assess each dimension required in a different random order.

Each delineated requirement was addressed in the following ways:

1. *Judges should all have experience within the domain in question. The experience does not have to be the same, but they should have had enough experience to develop implicit criteria for the measured dimension within the domain.* Three judges were used. They were all music educators, and all had experience working musically with children in 1st, 2nd, and 3rd grades. Two had earned doctorates in music, and the other one was currently enrolled in a doctoral program in music education. One judge specialized in theory and composition, one judge specialized in vocal music, and one judge specialized in instrumental music. The experience of the judges indicated the presence of implicit criteria needed to rate the quality of children's improvisations.
2. *Judges should judge independently of one another.* All judges viewed the improvisations and gave ratings without conferring with each other.
3. *Judges should not be trained or given a set of criteria by which to judge.* The interpretation of the term quality was left to the discretion of each judge. The instructions to the judges can be found in Appendix D.
4. *Judges should be required to make judgments on other aspects besides the targeted dimension, and instructed to keep these dimensions separate in their*

mind. This was the one instruction for implementation of the rating scheme that was discarded for this study for several reasons. In further reading of Amabile's (1983) work, she stated that providing ratings on other aspects would allow the researcher to see if these aspects were a part of what is deemed creativity, or separate dimensions in and of themselves. This would add to the body of research on the constructs of creativity. In a study that rated children's improvisations, Hickey (1995) found that the 3 musical dimensions rated were highly correlated. The correlations were: creativity and craftsmanship, .78; creativity and aesthetic appeal, .87; and craftsmanship and aesthetic appeal, .92. This present study was concerned with rating children's improvisations on the holistic aspect of "quality". It was quite probable that ratings on other dimensions of the improvisations would be highly correlational with the holistic dimension of quality. There was no intention in this study to research the underlying constructs of quality, so the addition of another rating variable would only serve to help focus the judges' minds on "quality" as opposed to another dimension of the improvisations.

Each dimension rated would require another complete viewing of the taped improvisations. The videotapes of the students' improvisations were quite lengthy, requiring approximately 4 ½ hours to view. Amabile (1983) studied the effect of task length on the reliability of judge's ratings and found that requiring longer concentration from judges resulted in lower interjudge reliability. For this study, the cost in judge fatigue was weighed against the

addition of another rating variable, and it was decided to only have the judges rate the improvisations on the dimension of quality.

5. *The products should be rated relative to each other, and not against a set standard within the domain.* To provide a framework for scoring, each judge viewed a preliminary tape of 15 randomly selected improvisations before viewing the tapes to be scored to gain an understanding of the breadth of quality present in 1st, 2nd, and 3rd grade children's improvisations.
6. *Each judge should view the products in a different random order, and assess each dimension required in a different random order.* A separate set of tapes was made for each judge that presented the improvisations in a unique random order. Because only one dimension was rated, the judges only assessed the improvisations once.

In summary, three independent judges were asked to rate each student's improvisations via a 5-point Likert scale on the aspect of quality. The interpretation of the term quality was left to the discretion of each judge. To provide the judges a framework for scoring as recommended by Amabile (1983), each judge viewed a preliminary tape of randomly selected improvisations before viewing the tapes to be scored. This tape contained 15 improvisations randomly selected via a randomly generated table. After viewing the preliminary tape, the judges then viewed the improvisations to be rated, which were presented in a unique random order to each judge, and gave each a rating on quality on a 5-point Likert scale. The judges were asked to use the entire Likert scale to provide differentiation between improvisations (see Appendix D

for a copy of the Instructions to Judges and Rating Scale). To provide a more complete understanding of the judges' ratings, descriptive information on examples of improvisations receiving each judge's highest and lowest scores can be found in Appendix E.

Reliability Estimates

The reliability estimates as determined by a Pearson product-moment correlation between judges were from .66 to .68 for the pre-test and from .58 to .64 for the post-test.

Table 2

Reliability Estimates for Judges' Pretest Ratings

| | Judge 1 | Judge 2 | Judge 3 |
|---------|---------|---------|---------|
| Judge 1 | 1.00 | .66 | .66 |
| Judge 2 | | 1.00 | .68 |
| Judge 3 | | | 1.00 |

Table 3

Reliability Estimates for Judges' Post test Ratings

| | Judge 1 | Judge 2 | Judge 3 |
|---------|---------|---------|---------|
| Judge 1 | 1.00 | .59 | .64 |
| Judge 2 | | 1.00 | .58 |
| Judge 3 | | | 1.00 |

The reliability estimates were lower than desired, but showed some agreement between judges. This analysis implies that each judge was rating differently—

using a differing paradigm by which to rate the improvisations. The average of the judges' ratings were used to analyze the differences between control and treatment groups. However, to fully understand the ratings of the judges because of the low reliability estimates, the ratings were also analyzed for significance by looking at each judge's individual ratings, as well as by looking at the mean ratings (see Appendix F for individual student's ratings). The change in improvisational quality as evidenced by the difference in students' ratings between pre and post test was subjected to an analysis of variance for the main effects of group (free play and non-free play) and grade to determine if there was a significant difference in improvisational quality between the free play and non-free play groups, or between grade levels for each of four data sets: mean ratings, and each of the three individual judge's ratings. A two-tailed test with a significance level of .05 was used.

CHAPTER IV

RESULTS

This study sought to discover the effect of musical free play experiences on the improvisational quality of first, second, and third grade children. Specifically, the questions were:

1. Is there a significant difference in the quality of musical improvisation of 1st, 2nd, and 3rd grade children receiving free play experiences during music instruction and the quality of musical improvisation of 1st, 2nd, and 3rd grade children not receiving free play experiences during music instruction?
2. Is there an interaction between grade level and the effects of free play on the quality of musical improvisation of 1st, 2nd, and 3rd grade children?

The study began with 110 students. Students were randomly divided into treatment and control groups. The final data was computed on 108 students because 2 students moved away from the school district during the study. Other than the treatment sessions, all other music instruction remained the same. The treatment took place over a period of 14 weeks. Both groups were tested with a researcher-designed instrument to elicit improvisations before and after treatment (see Appendix B). The improvisations were taped and rated by 3 independent judges on quality via a 5-point Likert scale. The judges were unaware of the grade level of the student, whether the improvisations were

from the control or treatment group, or whether the improvisations were elicited during the pretest or the post test. The determination of what comprised quality was left to each expert judge's discretion (see Appendix D for Instructions to Judges and Rating Forms).

Descriptive Data

Each improvisation received a rating from each of the three judges, and from these ratings a mean rating was calculated for each improvisation (see Appendix F for individual subject's ratings). From these individual mean scores the group mean, standard deviation, and minimum and maximum of the students' ratings were computed.

Table 4

Group Mean, Standard Deviation, Minimum, and Maximum of Students' Mean Ratings

| | Pretest | | | | Post test | | | |
|--------------|---------|-----|------|------|-----------|-----|------|------|
| | Mean | SD | Min | Max | Mean | SD | Min | Max |
| First grade | | | | | | | | |
| Control | 2.43 | .71 | 1.50 | 4.50 | 2.67 | .90 | 1.33 | 4.33 |
| Treatment | 2.22 | .69 | .83 | 3.17 | 2.34 | .66 | 1.00 | 3.33 |
| Total | 2.33 | .70 | .83 | 4.50 | 2.52 | .80 | 1.00 | 4.33 |
| Second grade | | | | | | | | |
| Control | 2.50 | .68 | 1.33 | 4.00 | 2.74 | .75 | 1.17 | 4.00 |
| Treatment | 2.61 | .81 | 1.17 | 4.17 | 2.66 | .49 | 1.83 | 3.83 |
| Total | 2.55 | .74 | 1.17 | 4.17 | 2.70 | .63 | 1.17 | 4.00 |

(table continues)

| | Pretest | | | | Post test | | | |
|-----------------|---------|-----|------|------|-----------|-----|------|------|
| | Mean | SD | Min | Max | Mean | SD | Min | Max |
| Third grade | | | | | | | | |
| Control | 2.73 | .56 | 1.50 | 3.83 | 2.68 | .49 | 1.67 | 3.50 |
| Treatment | 2.67 | .64 | 1.17 | 3.67 | 2.67 | .63 | 1.33 | 3.83 |
| Total | 2.70 | .60 | 1.17 | 3.83 | 2.68 | .56 | 1.33 | 3.50 |
| Control Total | 2.56 | .65 | 1.33 | 4.50 | 2.70 | .71 | 1.17 | 4.33 |
| Treatment Total | 2.52 | .73 | .83 | 4.17 | 2.58 | .60 | 1.00 | 3.83 |
| Total | 2.54 | .69 | .83 | 4.50 | 2.64 | .66 | 1.00 | 4.33 |

The judges' mean ratings ranged from .83 to 4.5, indicating that they had made use of the entire 5-point Likert scale in rating the improvisations as requested. The first grade contained the lowest and the highest ratings for both the pretest (.83, 4.50) and the post test (1.00, 4.33). Third grade showed the smallest range of ratings with a minimum of 1.17 and a maximum of 3.83. The minimum rating for the second grade was 1.17 and the maximum rating was 4.17. The first grade, when viewed as a total group, also showed the greatest variability between scores with a standard deviation of .70 for the pretest and .80 for the post test when compared to the other grade levels. The third grade group exhibited the smallest variability between scores with standard deviations of .60 for the pretest and .56 for the post test. Standard deviations for the second grade group were .74 for the pretest and .63 for the post test. When the groups were further divided into control and treatment subsets, the first grade again exhibited the greatest variance with a standard deviation of .90 exhibited in the post test ratings of the control group. The smallest

standard deviation of .49 was shared by the post test ratings of both the second grade treatment group and the third grade control group.

When all three grades were considered together, the means for the pretest for both the control and treatment groups were remarkably close together. The mean of the control group was slightly higher (2.56) than that of the treatment group (2.52). The means for the post test were further apart with the control group (2.70) still being higher than the treatment group (2.52). The control group exhibited higher minimums and maximums (1.33, 4.50) than did the treatment group (.83, 4.17). The standard deviations for both groups were similar. The control group displayed standard deviations of .65 and .71 for the pretest and the post test respectively. The standard deviations for the treatment group were .73 and .60 for the pretest and the post test respectively. The standard deviations were small, indicating that even though the judges made use of the entire Likert scale in rating the improvisations, there was not a large variance in the students' performance as displayed by the ratings.

The means for the control groups in both the first and second grades showed the largest growth between pretest and post test with a change in ratings of .24. The first grade treatment group showed a change in ratings between pretest and post test of .12. The second grade treatment group showed a change in ratings between pretest and post test of only .05. The third grade ratings showed a non-changing mean rating of 2.67 between pre and post test for the treatment group and a negative change in ratings between pre and post test (-.15) for the control group.

This descriptive data shows that the greater variability between scores, and the subsequent greater change in scores was seen in the younger students. The third grade showed less variability in scores, and a static state or negative measured growth between the pretest and the post test ratings. There was a small change in the mean ratings between grade levels visible in the pretest ratings, with each grade level getting progressively higher ratings (first grade, 2.33; second grade, 2.55; and third grade, 2.70). The post test ratings, however, showed a growth in the mean ratings for the first and second grades, and a slight decline in the mean rating for the third grade (first grade, 2.52; second grade, 2.70; and third grade, 2.68).

Pretest and Post Test Histograms

To further understand the distribution of scores, histograms were constructed representing pretest and post test ratings for each grade level, as well as for the total group. A normal curve has been superimposed onto each histogram.

The histograms show distributions that are somewhat normally distributed, with small standard deviations in evidence. All distributions have large concentrations of scores around the mid-range (2.0-3.0). The first grade pretest (Figure 1) shows a large grouping of students in the middle and lower half of the rating scale. The first grade post test (Figure 2) shows a mild drop in the lowest ratings, and a slight gain in the higher ratings, but with the majority of the ratings still at or below the mid-range. The second grade also shows a large grouping in the mid-range (Figures 3 and 4). The pretest also shows a shape approximating a normal curve. The post test shows a few more students scoring in the upper half, but still a preponderance of ratings in the mid-range. The third

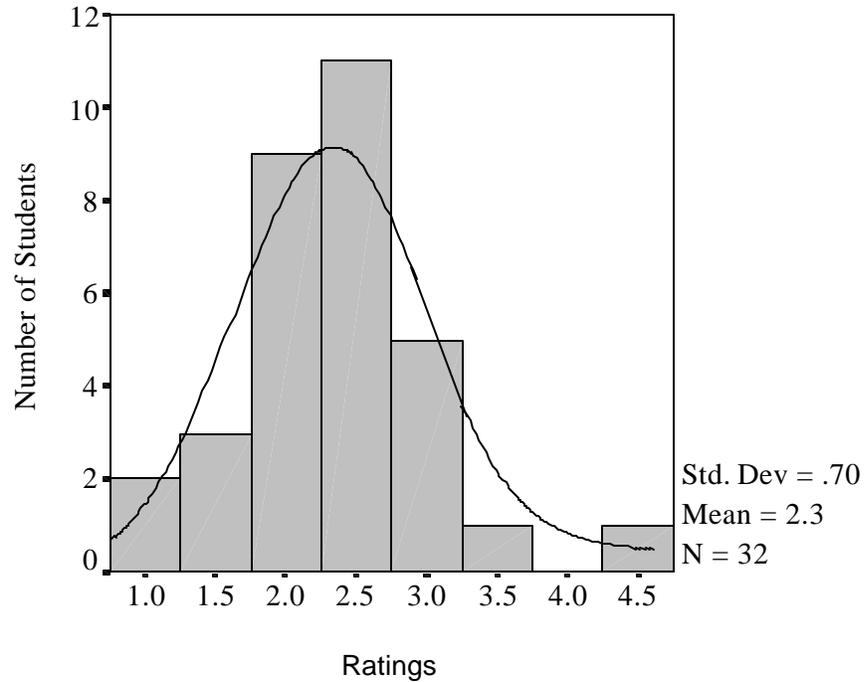


Figure 1. Histogram of mean ratings of 1st grade students' pretest improvisations

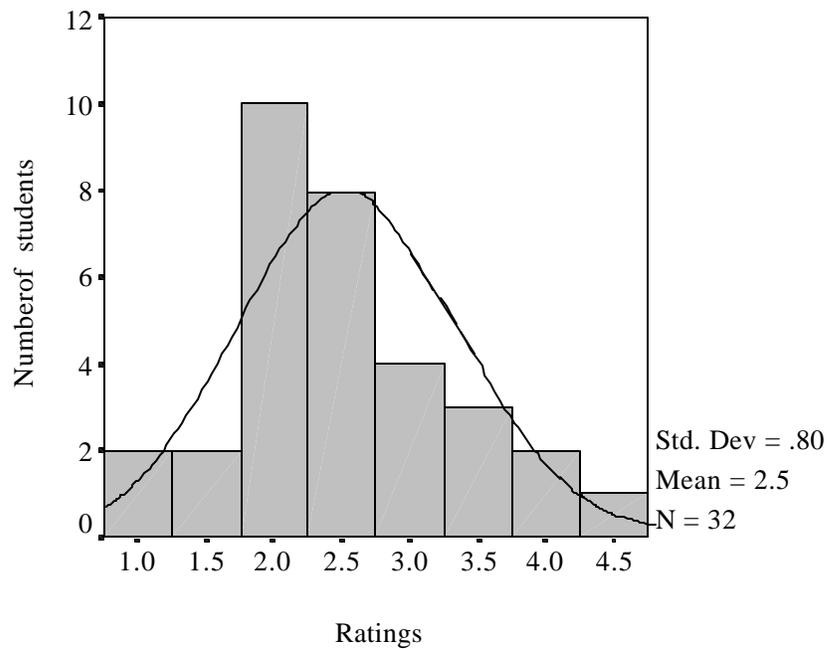


Figure 2. Histogram of mean ratings of 1st grade students' post test improvisations.

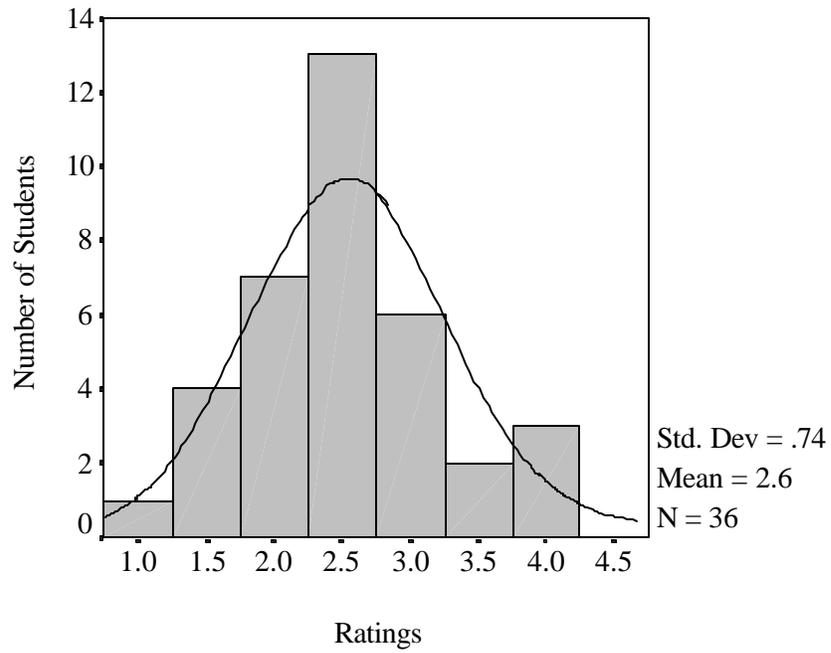


Figure 3. Histogram of mean ratings of 2nd grade students' pretest improvisations.

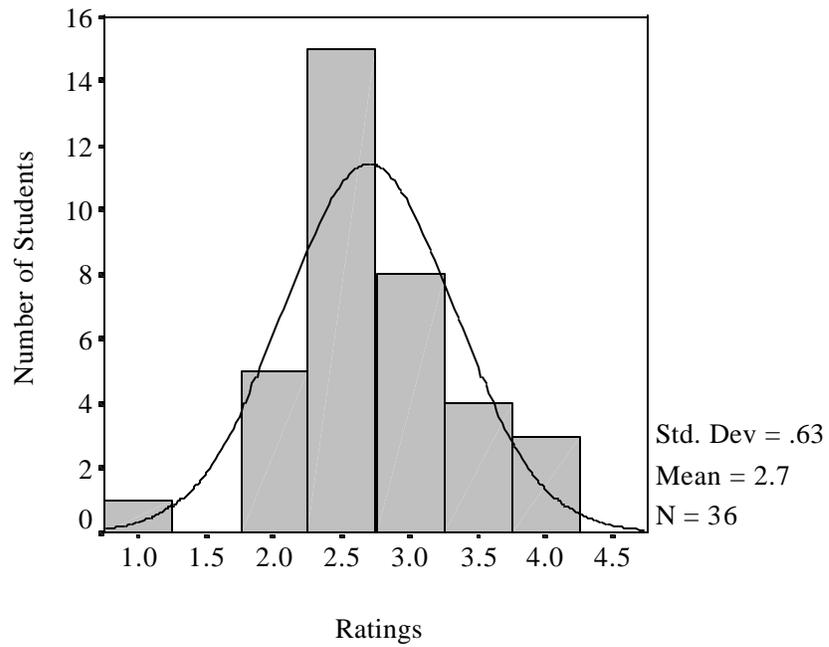


Figure 4. Histogram of mean ratings of 2nd grade students' post test improvisations.

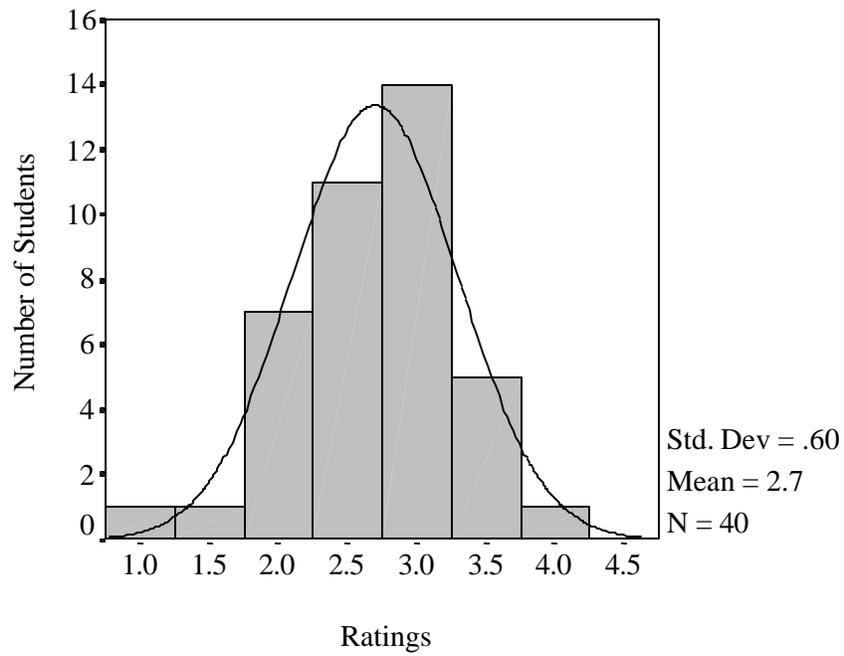


Figure 5. Histogram of mean ratings of 3rd grade students' pretest improvisations

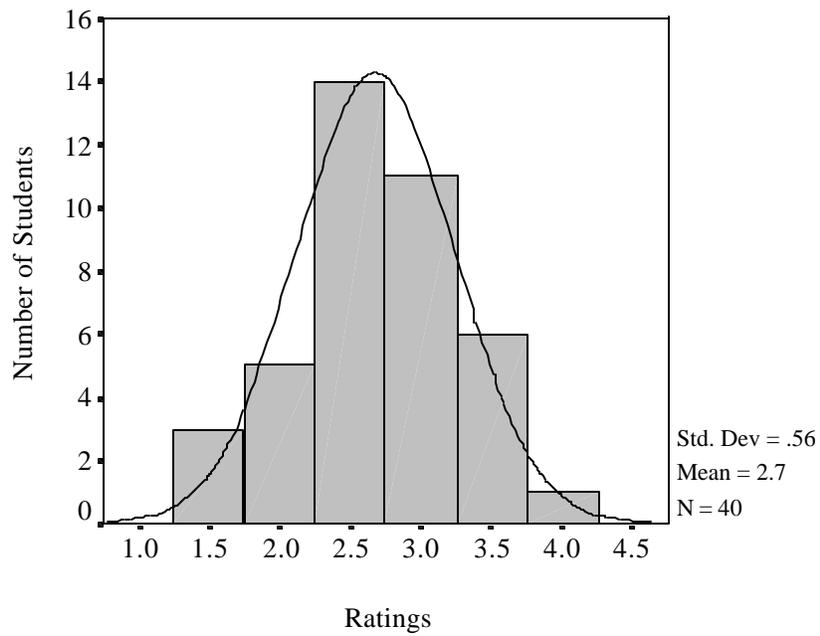


Figure 6. Histogram of mean ratings of 3rd grade students' post test improvisations

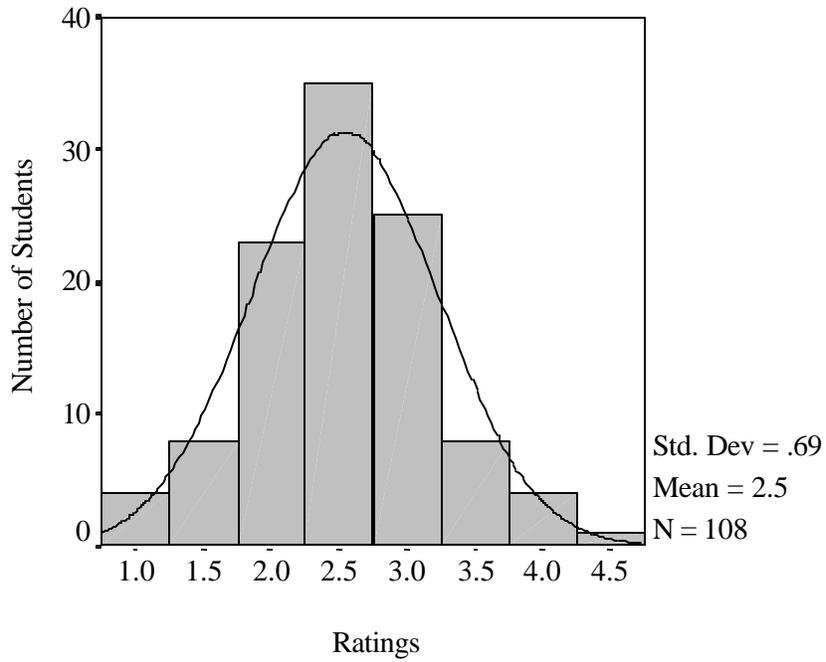


Figure 7. Histogram of mean ratings of total group's pretest improvisations

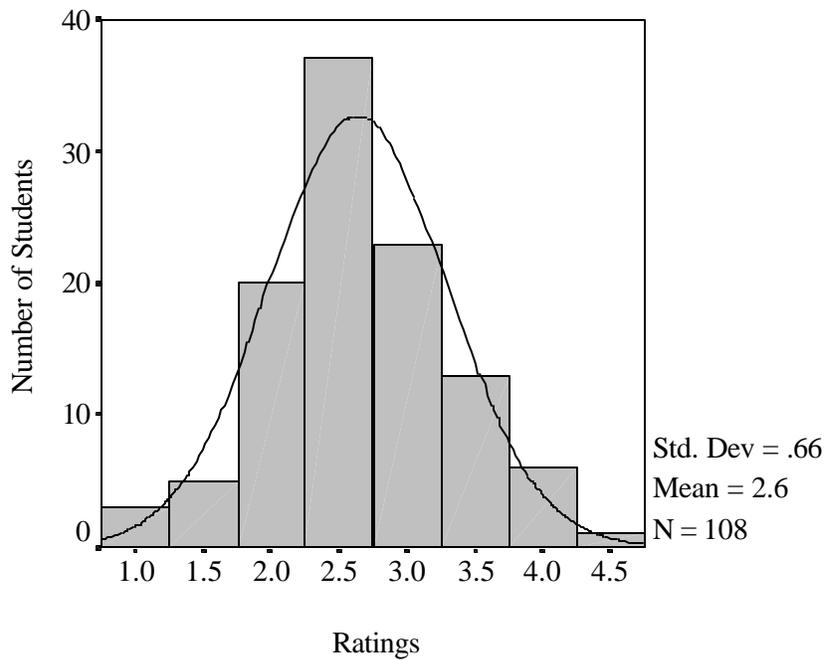


Figure 8. Histogram of mean ratings of total group's post test improvisations

grade distributions show a shift of concentration of students, while retaining the same mean score of 2.7. The mode changes from 3.0 in the pretest to 2.5 in the post test (Figures 5 and 6). When groups were considered as a whole (Figures 7 and 8), they exhibited similar patterns for both pretest and post test ratings. Both the overall pretest and overall post test ratings were very close to a normal curve, with larger groupings around the mid-range. The histograms show a normal distribution of ratings over the total group, with some abnormalities seen within smaller sub-groupings.

Determination of Statistical Significance

To determine whether the differences between groups were significant, an analysis of variance (ANOVA) was utilized. However, before an ANOVA was undertaken, the ratings for the control and treatment groups were compared for homogeneity of variance.

Homogeneity of Variance

One of the assumptions underlying ANOVA is that of homogeneity of variance (Hinkle, Wiersma, and Jurs, 1988). Groups being compared must exhibit variances between scores that are similar enough to be representative of the same population to produce robust findings. According to Hinkle, Wiersma, and Jurs (1988), if the sample sizes are equal, unequal variances have minimal impact on the resulting ANOVA statistic. If, however, the sample sizes are unequal, the issue of homogeneity of variance becomes more important. If the larger sample size exhibits a larger variance, the F test will be too conservative. In other words, if the α level is .05 the actual level of significance will be less than .05. If the larger sample size exhibits a smaller variance, the

F test will be too liberal. This study experienced unequal sample sizes because two 1st grade students moved away during the study. This left 55 students in the total control group, and 53 students in the total treatment group. Because of unequal sample sizes in the first grade in this research project, an F-test for homogeneity of variance was computed for both the total group and for the first grade.

Table 5

F-test for Homogeneity of Variance

| | Variance Control Group | Variance Treatment Group | Observed F | F _{cv} ($\alpha=.05$) |
|-------------|---------------------------|-----------------------------|------------|-------------------------------------|
| First Grade | | | | 2.445 |
| Pretest | 0.500 | 0.475 | 1.053 | |
| Post test | 0.806 | 0.432 | 1.865 | |
| Total Group | | | | 1.580 |
| Pretest | 0.421 | 0.531 | 1.261 | |
| Post test | 0.503 | 0.361 | 1.394 | |

The observed F statistics for the first grade were 1.053 for the pretest and 1.865 for the post test, which were below the critical value of 2.445. The observed F statistics for the total group were 1.261 for the pretest and 1.394 for the post test, which were below the critical value of 1.580. Because observed F statistics were below the corresponding critical values, variances were determined to be homogeneous between groups at an alpha level of .05. Therefore, with homogeneity of variance established for

total group and the first grade subgroup, an analysis of variance could then be computed even though the study contained unequal sample sizes.

Analysis of Variance

The research questions asked whether free play experiences had a significant effect on the improvisational quality of the students, and whether grade level was a factor in any effect of free play on improvisational quality. To determine this, an analysis of variance was performed ($\alpha=.05$) on the difference in ratings between the pretest and post test for the main effects of group (control and treatment) and grade level.

The reliability estimates between judges were from .66 to .68 for the pre-test and from .58 to .64 for the post-test. These estimates could indicate that the judges were using differing criteria for quality by which to rate the improvisations, and thus could provide differing values of measured change in improvisational quality. Therefore, an ANOVA was computed for each of four data sets—mean ratings and each of the three judges' ratings—to determine if each individual judge's ratings showed the same findings as the pooled ratings.

Analysis of Mean Ratings

An analysis of variance was computed on the mean ratings of the three judges to determine if statistical differences existed in the change in ratings between the pretest and post test for the main effects of group (control and treatment) and grade level.

Table 6

Mean Ratings—Analysis of Variance (Difference by Group, Grade)

| | | Sum of Squares | df | Mean Square | F | Sig. $\alpha=.05$ |
|--------------------|-----------------|----------------|-----|-------------|-------|-------------------|
| Main Effects | Group | .151 | 1 | .151 | .359 | .550 |
| | Grade | .923 | 2 | .462 | 1.101 | .337 |
| 2-way Interactions | Group and Grade | .324 | 2 | .162 | .386 | .681 |
| Model | | 1.394 | 5 | .279 | .665 | .651 |
| Residual | | 42.779 | 102 | .419 | | |
| Total | | 44.173 | 107 | | | |

In Table 9 the two groups were control and treatment. There were 55 students in the control group and 53 students in the treatment group. The analysis of variance table shows that there is not a statistically significant difference at an alpha level of .05 in the change in pretest and post test ratings between control and treatment groups (observed F .359, p-value = .550). The table also shows that there is not a statistically significant difference at an alpha level of .05 in the change in pretest and post test ratings for grade levels (observed F = 1.101, p-value = .337). There is also not a statistically significant interaction at an alpha level of .05 demonstrated between the variables of group and grade (observed F = .386, p-value = .651). The analysis of the mean ratings data set reveals no significant difference in change of ratings between control and treatment groups for the group as a whole, or for any particular grade level within the total group.

Analysis of Individual Judge's Ratings

Each individual judge's ratings was subjected to an analysis of variance to determine if the judges, each using his or her self-defined criteria for quality, produced findings that supported those shown in the analysis of the mean ratings.

Table 7

Judge One's Ratings—Analysis of Variance (Difference by Group, Grade)

| | | Sum of Squares | df | Mean Square | F | Sig. $\alpha=.05$ |
|--------------------|-----------------|----------------|-----|-------------|-------|-------------------|
| Main Effects | Group | .412 | 1 | .412 | .914 | .341 |
| | Grade | 1.134 | 2 | .567 | 1.257 | .289 |
| 2-way Interactions | Group and Grade | .163 | 2 | 8.171E-02 | .181 | .835 |
| Model | | 1.687 | 5 | .337 | .748 | .589 |
| Residual | | 45.995 | 102 | .451 | | |
| Total | | 47.682 | 107 | | | |

The ANOVA table for Judge One shows that there is not a statistically significant difference at an alpha level of .05 in the change in pretest and post test ratings between control and treatment groups (observed F .914, p-value = .341). The table also shows that there is not a statistically significant difference at an alpha level of .05 in the change in pretest and post test ratings for grade levels (observed F = 1.257, p-value = .289). There is also not a statistically significant interaction at an alpha level of .05 demonstrated between the variables of group and grade (observed F = .181, p-value = .835). An analysis of Judge One's ratings reveals no significant difference in change of ratings

between control and treatment groups for the group as a whole, or for any particular grade level within the total group. This supports the findings of the mean ratings data set that established no significance between control and treatment groups for the group as a whole, or for any particular grade level within the total group.

Table 8

Judge Two's Ratings—Analysis of Variance (Difference by Group, Grade)

| | | Sum of Squares | df | Mean Square | F | Sig. $\alpha=.05$ |
|--------------------|-----------------|----------------|-----|-------------|-------|-------------------|
| Main Effects | Group | .501 | 1 | .501 | .559 | .456 |
| | Grade | 1.840 | 2 | .920 | 1.027 | .362 |
| 2-way Interactions | Group and Grade | .280 | 2 | .140 | .156 | .856 |
| Model | | 2.675 | 5 | .535 | .597 | .702 |
| Residual | | 91.351 | 102 | .896 | | |
| Total | | 94.025 | 107 | | | |

The ANOVA table for Judge Two shows that there is not a statistically significant difference at an alpha level of .05 in the change in pretest and post test ratings between control and treatment groups (observed F .559, p-value = .456). The table also shows that there is not a statistically significant difference at an alpha level of .05 in the change in pretest and post test ratings for grade levels (observed F = 1.027, p-value = .362). There is also not a statistically significant interaction at an alpha level of .05 demonstrated between the variables of group and grade (observed F = .156, p-value = .856). An analysis of Judge Two's ratings reveals no significant difference in change of ratings

between control and treatment groups for the group as a whole, or for any particular grade level within the total group. This supports the findings of the mean ratings data set that established no significance between control and treatment groups for the group as a whole, or for any particular grade level within the total group.

Table 9

Judge Three's Ratings—Analysis of Variance (Difference by Group, Grade)

| | | Sum of Squares | df | Mean Square | F | Sig. $\alpha=.05$ |
|--------------------|-----------------|----------------|-----|-------------|-------|-------------------|
| Main Effects | Group | 3.455E-02 | 1 | 3.455E-02 | .048 | .828 |
| | Grade | 1.963 | 2 | .982 | 1.352 | .263 |
| 2-way Interactions | Group and Grade | 1.235 | 2 | .617 | .850 | .430 |
| Model | | 3.230 | 5 | .646 | .890 | .491 |
| Residual | | 74.038 | 102 | .726 | | |
| Total | | 77.269 | 107 | | | |

The ANOVA table for Judge Three shows that there is not a statistically significant difference at an alpha level of .05 in the change in pretest and post test ratings between control and treatment groups (observed F .048, p-value = .828). The table also shows that there is not a statistically significant difference at an alpha level of .05 in the change in pretest and post test ratings for grade levels (observed F = 1.352, p-value = .263). There is also not a statistically significant interaction at an alpha level of .05 demonstrated between the variables of group and grade (observed F = .850, p-value = .430). An analysis of Judge Three's ratings reveals no significant difference in change of

ratings between control and treatment groups for the group as a whole, or for any particular grade level within the total group. This supports the findings of the mean ratings data set that established no significance between control and treatment groups for the group as a whole, or for any particular grade level within the total group.

No significant difference was found by any individual judge, or by the judges as a whole, in the change in ratings between control and treatment groups for the group as a whole, or for any particular grade level within the total group.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS FOR FURTHER STUDY

Summary

This study sought to discover the effect of musical free play experiences on the improvisational quality of first, second, and third grade children. Specifically, the questions were:

1. Is there a significant difference in the quality of musical improvisation of 1st, 2nd, and 3rd grade children receiving free play experiences during music instruction and the quality of musical improvisation of 1st, 2nd, and 3rd grade children not receiving free play experiences during music instruction?
2. Is there an interaction between grade level and the effects of free play on the quality of musical improvisation of 1st, 2nd, and 3rd grade children?

A pretest post test control group design was chosen for this study. There were 55 students in the control group and 53 students in the treatment group. All subjects were tested with a researcher-designed instrument to elicit improvisatory behavior (see Appendix B). After testing, the free play treatment was administered in seven different treatment sessions to the treatment group over a period of 14 weeks. During the same time period, the control group received regular music instruction. Treatment and

control groups also participated in normally scheduled music classes during these 14 weeks. At the end of the 14 weeks, all subjects were again tested using the same researcher designed instrument to elicit improvisatory behavior. The taped improvisations were then submitted to three independent judges. The judges were unaware of the grade level of the student, whether the improvisations were from the control or treatment group, or whether the improvisations were elicited during the pretest or the post test. The determination of what comprised quality was left to each expert judge's discretion. The improvisations were rated on the aspect of quality on a 5-point Likert scale.

The change in ratings between the pretest and post test for the two groups was then subjected to an analysis of variance to determine whether the free play experiences had a significant impact on the quality of the subjects' improvisations. The analysis of variance on the mean ratings showed no significant difference between control and treatment groups for the group as a whole, or for any of the grade level groupings. Because of lower than desired interjudge reliability estimates, which ranged from .58 to .68, an analysis of variance was computed for each individual set of ratings from the three judges. The separate analysis of variance computed for each of the three judges supported the findings of the analysis of the mean ratings—no significant differences were found between any groups.

Findings

In response to research question one, there is no significant difference in the quality of musical improvisation of 1st, 2nd, and 3rd grade children receiving free play

experiences during music instruction and the quality of musical improvisation of 1st, 2nd, and 3rd grade children not receiving free play experiences during music instruction.

In response to research question two, there is no interaction between grade level and the effects of free play on the quality of musical improvisation of 1st, 2nd, and 3rd grade children.

Research Design and Subsequent Findings

In developing the design of this study, decisions were made concerning research design that affect the interpretation of the findings presented. The aspects of sample size, small variability in the measurement instrument, interjudge reliability, and length of study will be discussed in relationship to the findings presented.

Sample Size and Variability of the Measurement Instrument

The sample size for this study was large enough to detect an effect size of .75 with the power of the test being .90 at an α of .05 for comparison between the total play and non-play groups. However, to compare the interaction of grade on any differences between groups, the sample size recommended in each of 6 groups (2 treatments X 3 grade levels) at the same level would be 52. This would have resulted in a total sample size of 312 students. In developing the design of this study, the sample size was kept relatively small ($n = 108$) to limit judge fatigue and to allow implementation of the treatment at one school setting to avoid introducing extraneous variables into the research. According to Hinkle, Wiersma, and Jurs (1988), a smaller sample size may prevent small differences between groups from being perceptible—which may have been the case when looking at the differences between grade levels. In addition, the rating

scale was kept to a 5-point limit to help reduce variability between judges. The 5-point limit required a larger difference between groups before any significant difference could be determined. These aspects in the design of this study make it vulnerable to a Type II error. However, the descriptive data showed that in the first and second grades, the growth in improvisational quality between the pretest and the post test, even though not at a level of significance, was greater in the control group than in the treatment group. At the third grade level, the mean of the treatment group remained constant, while that of the control group fell—but only slightly. When the group was considered as a whole, the growth in improvisational quality between the pretest and the post test, even though not at a level of significance, was again greater in the control group than in the treatment group. Considering the actual mean ratings of the students' improvisations, it is unlikely that the smaller sample size or small variability in the Likert scale caused true differences between groups to remain hidden. Thus, the findings of no positive effect of play experiences on the quality of children's improvisation appears to be justifiable.

Interjudge Reliability

The measurement protocol was modeled after Amabile's consensual assessment technique (Amabile, 1983). As a part of this technique, expert judges do not undergo any special training in the rating of the target variable, but use their own criteria developed from experience in the domain in question. This produces a rating that is similar to "real world" ratings of subjective concepts such as quality (Amabile, 1983). However, this measurement technique produced interjudge reliability estimates that ranged from .58 to .68—lower than desired to produce robust findings in a research project. Further analysis

of the judges' ratings, however, showed that each judge varied consistently with the other judges. This finding supported the use of the means of the judges' ratings for the final analysis, as recommended by Amabile (1983). To further substantiate the findings, an analysis of variance was calculated for each of the individual judge's ratings. The analysis of the ratings from each judge supported the analysis of the mean ratings—there was no significant difference between groups. This suggests that the finding of no significant difference between groups as to the variable of quality is robust whether quality is viewed from each individual expert judge's perspective, or from a perspective that takes into account the combined measurement paradigms employed by the three judges.

Length of Study

The length of time chosen for this study was 14 weeks. The literature relating to play's effect on children's development was mixed as to the length of time required to ascertain any benefit. Some studies reported a positive effect after one to three 10-minute sessions (Dansky & Silverman, 1973, 1975; Li, 1978; Pepler & Ross, 1981; Smith & Dutton, 1979; Smith & Simon, 1983; Smith, Simon & Emberton, 1985; Vandenberg, 1981) Other studies described benefits received from play that took place over a period of several years (Curran, 1993; Hartmann & Rollett, 1994; and Sefer, 1995). The current study, which lasted 14 weeks, fell in-between the extremes as to its longevity. The length of 14 weeks was chosen because it was felt that this provided enough time to render discernible results, while keeping the project within a semester's limit. It is possible that a longer play treatment would have produced an effect that was not produced by this study.

The possible role of length of play treatment in the research setting needs to be investigated more extensively. Another complicating factor may have been the interaction of research length, adult intervention, and type of variable chosen for study. These issues will be explored more in-depth in the next section.

Conclusions

The main conclusion drawn from this study is that free play was not beneficial, under the specific conditions of this study, for the improvement of children's improvisational quality. How does this finding fit in to the body of literature on which this study was based? The discussion of this can be divided into three sections: studies on play and its effect on cognitive development, studies on play and its effect on musical development, and issues of measurement.

Play and Cognitive Development

A comparison of this study with previous cognitive development studies reveals both similarities and differences in treatments and measurement tasks that may have accounted for the contradictory findings. These similarities and differences will be explored before conclusions are drawn.

Treatment Considerations

The treatments of these various studies contained important differences in the length of the treatment, location of play treatment, and amount of adult intervention present in the treatment. Nine studies reviewed (Dansky & Silverman, 1973, 1975; Li, 1978; Pepler & Ross, 1981; Smith & Dutton, 1979; Smith & Simon, 1983, 1985; Smith, Simon & Emborton, 1985; Vandenberg, 1981) confined their length of treatment to one

to three 10-minute sessions. Three studies reviewed looked at the effect of play distributed over years (Curran, 1993; Hartmann & Rollett, 1994; and Sefer, 1995).

Of the studies reviewed utilizing brief play treatments, only one (Smith & Simon, 1985) produced no significant effect resulting from the play treatment. All of the other studies showed varying degrees of positive effect from participating in play. The study by Smith and Simon (1985) showing no significant effect was, however, the study with the most stringent research controls—implementing controls for testing bias, scoring bias, and treatment bias that the researchers believed had plagued previous play research. Research methods attempting to control for the biases highlighted by Smith and Simon were implemented in this study. After designing the study to help control for these biases, the results were the same.

The studies looking at longer play treatments all found significant positive effects from the play treatment. The current study, which found no positive effects from the play treatment, was 14 weeks long, much longer than those studies examining 10 minutes of play, but much shorter than those studies that examined effects after years of play experiences.

Of the brief studies, all but one (Li, 1978) studied play without adult intervention. Li (1978) used minimal adult intervention in the attempt to provide a pretend mindset in the child before play. All of the brief studies, with the exception of the 1985 Smith and Simon study saw positive results for play—either with minimal or no adult intervention. The studies that looked at longer periods of play (Curran, 1993; Hartmann & Rollett, 1994; and Sefer, 1995) used a mixture of adult intervention and free play. The amount of

adult intervention in the play activities was not a specified part of the treatment, nor was it measured, however—which was problematic because it was impossible to partition off which part of the measured gain was due to play and which part was due to specific adult intervention. The current study sought to minimize adult intervention, to explore the effect of free play, as did the brief studies—but found no gains in improvisational quality because of participation in the play experiences.

Task Considerations

The measurement tasks could be divided into two kinds: those that relied on objective measurement, and those that relied on subjective measurement. The brief studies relied mostly on objective measurement of distinct behaviors. These measurements included counting how many uses a child can name for a particular object, the amount of time needed to solve a problem, the number and level of hints needed to solve a problem, or how much time was spent on and off task. Some of these objective studies categorized answers into standard and non-standard uses for objects, which became somewhat subjective, but interrater agreement was quite high as to what constituted standard and non-standard uses. At no time did the brief studies attempt to measure a holistic aspect such as quality.

The study by Curran (1993) looked at children's play habits and used a subjective measure to divide the children into categories as to their play habits. The effects of these play habits, however were measured with several objective measures. The two longitudinal studies that exposed children to play treatments (Hartmann & Rollett, 1994;

Sefer, 1995) used subjective measures of holistic behavior to provide a basis by which to compare groups of students.

The current study utilized a subjective measure to measure holistic behavior by which to compare groups, as did the Hartmann and Rollett (1994) and Sefer (1995) studies. The measurement technique utilized was similar to these two longitudinal studies—it compared the quality of task completion between groups by means of expert judges after one group has received a series of play experiences. However, this study was implemented with a 14-week treatment period, while the longitudinal studies looked at play taking place over a period of 4 to 5 years. It is possible that any benefits to holistic aspects of children’s cognitive growth that are derived from play can only be seen after experiencing play over a longer period.

Conclusions

The findings of this study contradict the findings of the majority of studies reviewed in the area of play and cognitive development. Reasons for this might be that participating in shorter play treatments may produce significant differences in discrete behaviors in the short term, whereas any discernible differences in the total quality of work may not be visible in anything less than long-term studies. The conclusion of the Simon and Smith (1985) who undertook the study that also produced no positive effects after play treatment, was that one 10-minute session is not enough time for benefits of play to be realized. They suggested longer periods of time to fully appreciate the benefits of play. This study, consisting of seven 20-minute play sessions over a period of 14

weeks, may not have been long enough to produce results in holistic behaviors. The question of how long is long enough is still unanswered.

Play experiences may also need to have adult intervention to produce the holistic changes that were discernible in the Hartmann and Rollett (1994) and Sefer (1995) studies. The current study was designed to remove adult intervention as a variable, which may have contributed to the lack of discernible change. It is also possible that the interaction of length of play and amount of adult intervention could have produced the contradictory findings seen when the literature is viewed as a whole.

This study has provided a point of incongruity with the body of literature in the area of play's effect on cognitive development. The salient points that could have caused this incongruity seem to be length of study, type of dependent variable, adult intervention, or an interaction of these variables.

Play and Music

The writings that provided support for the premise that play would enhance improvisation in children (Addison, 1991; Ensley, 1987; and Swanwick, 1988) compared the process of improvising or composing music to the process of play, and found them similar. These writers built a strong argument for the inclusion of play in music instruction. Studies carried out in the field by researchers (Kartomi, 1991; Marsh, 1995; Merrill-Mirsky, 1988; and Riddell, 1990) documented children participating in musical activities on the playground. The conclusions drawn from these researchers were that the play aspect of playground music performance and the subsequent peer teaching evolving from this play should be incorporated into the music classroom. Other studies reviewed

(Littleton, 1991; Miller, 1983; Moorhead & Pond, 1978; Shelley, 1981; and Veldhuis, 1992) documented behaviors during play in a classroom setting that pointed to the educational benefits of free play within the music curriculum. In the informal observations of children during their free play experiences during this study, the children were found to participate in the same behaviors observed in previous studies on musical play. Activities in which the students engaged included instrument and sound exploration, figuring out songs by ear, singing, moving, and making up music alone and in groups. Some students spent time pretending to teach each other in a music class, taking turns being the teacher. Some students dressed up wearing as many instruments as they could and then danced to make music. The students also organized concert bands, marching bands, and rock bands with the appropriate conductors, baton twirlers, or lead singers. Several children linked together the headsets and microphones for the 3 portable pianos available to them and pretended to be airplane pilots making a variety of sound effects with the pianos and their voices. A major preoccupation that happened every free play session was drumming together in groups ranging in size from two to eight. This study, however, found that those same free play behaviors, carried out over a semester's time, did not significantly enhance the quality of children's improvisations.

The findings of this study suggest that free play during a semester's time, in and of itself, does not produce the growth in improvisational quality that was presumed by those observing musical play. A variable that may have affected the findings of this study as to effect of play on improvisational quality could be that of time. As discussed previously, a semester's worth of play may not have been enough to effect any

discernible changes in a holistic aspect such as quality of improvisation. The question then could be asked: “How much time is enough?” This study devoted 20 minutes out of every 2 hours available for instruction during a semester to free play. Is the amount of time needed a greater percentage of available time, or is the play time more valuable if experienced over a few years? This question is worthy of exploration. However, the answer may lie elsewhere—namely in the issue of adult intervention in play.

The play investigated in this study was a pure form of play—free play. The research was designed to limit adult intervention as much as possible to look at play, in and of itself. By removing adult intervention as much as possible, any measured results could be reasonably assured to be from the play and not from the finesse of any particular adult intervening in the play. This free play—without adult intervention—produced behaviors that corresponded to the behaviors other researchers had found. These behaviors such as sound exploration, group impromptu performance, pretend play involving music making, dancing, arranging instruments in groupings (such as low to high, or sameness of sound source), or making up songs would lead many educators to think that much learning was happening. However, these behaviors, as delightful as they may be, may not actually lead to growth in the quality of children’s improvisations. It may be that adult intervention in the play process is necessary for the observed play behaviors to contribute to the growth of improvisational quality. For growth in improvisational ability to occur, we may need to directly teach to that task. Merely being exposed to the improvisational materials, and being allowed to experiment with them may not lead to an improvement in quality of the child’s improvisations.

The findings of this study which contradict the writings of those teachers experiencing success with a music curriculum that includes play, underscores the need of those looking at play in music education to also explore the issue of adult intervention in play. At which point of adult intervention in children's play does the "play" cease to be play and become instruction? In supporting a music curriculum that includes play, are we actually advocating play as children experience play, or are we advocating another type of teaching style? If we are able to clearly delineate these boundaries, we will be able to integrate findings from observational studies of free play into our teaching strategies.

Issues of Measurement

This study used Amabile's consensual assessment technique (Amabile, 1983) to rate the quality of children's improvisations. The consensual assessment technique, in various modified forms, has been used to rate musical compositions in research situations by Bangs (1992), Brinkman (1994), Hickey (1995), and Webster and Hickey (1995). The studies undertaken by Bangs (1992), Brinkman (1994), and Webster and Hickey (1995) experienced interjudge reliability estimates at relatively high levels (.76 to .96). However, Hickey (1995) experienced unsatisfactory interjudge reliability estimates (.18 to .80) and chose to discard 1/3 of her data. The remaining data—which rated the highest and lowest compositions—produced reliability estimates ranging from .62 to .93, which she judged satisfactory.

In this study, initial interjudge reliability estimates ranged from .58 to .68, requiring further exploration before they could be used with confidence. When the judges' ratings were analyzed further, however, they were found to vary somewhat

consistently with each other. This supported the use of the means of the judges' ratings for ascertaining the change in improvisational quality between the pretest and the post test as called for by Amabile (1983). To further examine the findings, each judge's individual ratings were subjected to an analysis of variance to see if they agreed with the main findings. The findings from the analysis of variance of each individual judge's ratings supported the main finding of no significant difference between groups. Each judge, working within his or her own paradigm, had produced findings that substantiated the ratings as a whole.

The ultimate confidence in the findings achieved in this study supports the use of the consensual assessment technique. Use of the consensual assessment technique provides researchers with a measurement tool that approximates real world assessment, which is helpful in designing research that will benefit educators. However, the original reliability estimates produced in this study were disconcerting. In Amabile's work, she does caution against judge fatigue as it contributes to lower reliability estimates, which may have been the case in this study as each judge rated 216 improvisations. However, to be useful as a research tool, any rating technique utilized needs to be able to deal with larger sample sizes necessary to produce statistically significant results. The consensual assessment technique is of value for the holistic rating of musical composition and improvisation. However, the problem of low interjudge reliability needs to be addressed through modification of the technique to facilitate the use of this tool in research situations.

Recommendations for Further Study

One of the areas that appears to need more study is that of reconciling the differing views of just what constitutes musical play. Observational studies have focused on free play, while those writers providing description of music instruction incorporating musical play seem to incorporate varying amounts of adult intervention. Theoretical work that explored the definition of play outside of music and then applied these definitions to play taking place within music education would provide the researcher more precise measurement tools with which to work, and allow research on play to be shared across disciplines.

This study attempted to start from the bottom up—that is to look at a pure form of play that contained as little adult intervention as possible and to discover its effect on one aspect of music education—the quality of improvisation. Another way to look at this might be to work from the top down—that is to define and describe successful musical play experiences, and document the adult intervention involved in the implementation of them. Varying amounts of adult intervention in play could then be tested for effectiveness.

It is also possible that play contributes to other areas of musical growth. The literature reviewed points to the ability of play to enhance positive attitudes towards school, development of technical ability through repetition, and development of vocal abilities. Research could be carried out in each of these areas to determine the role of play in their development.

This study looked at improvisational quality—a holistic variable. Although there were no significant differences in the measured quality between groups, it is possible that the improvisations contained other differences. Specific aspects of children's improvisational work may be affected by their participation in play experiences. Research designed to understand how play experiences affect musical characteristics such as length, melodic repetition and development, use of rhythm and timbre, or the child's ability to replicate the improvisation would be of use in understanding the development of improvisational ability in children.

Children were observed participating in a variety of activities during their play time. A study that compared the activities in which the child engaged while playing and any subsequent changes in improvisational process or product might provide more insight into the role of play in this area of musical development. Information on which play activities were of more educational value would assist in designing adult intervention to make the most use of children's play.

There was no significant growth in improvisational quality found for any of the groups over the semester. This in and of itself was an interesting finding. It is possible that the eclectic curriculum pursued in this school is not conducive to growth in improvisational ability. In this study, no effort was made to modify the basic music curriculum already in place in the school. It would be enlightening to spend concentrated time teaching improvisation skills, with and without free play time allotted, to discern if free play, as part of a dedicated program in improvisation and composition, would enhance the quality of children's improvisations.

Personal Observations

As with most studies, this one provided peripheral information that was not a part of the original research questions. Other researchers pursuing a study of play and young children may find the following observations helpful. The observations have been divided into two sections: suggestions for the implementation of musical play and personal observations on musical play.

Suggestions for the Implementation of Musical Play

From implementing the treatment in two different schools, some procedural matters that may be of help to the classroom teacher wishing to try some play sessions, or the researcher wishing to look at play are offered here. These are divided into three sections: physical recommendations, procedural recommendations, and instrument considerations.

Physical Requirements

The room in which the play is experienced, needs to be as large as possible, with as much sound absorbing material in the carpets, walls, and ceilings as possible. Noise is problematic. The play area should also be located in a part of the school building in which it will not disturb others trying to have a “normal” classroom experience. Clocks should be visible, as bells won’t be able to be heard. Also, the administration needs to be aware that any bells rung for emergency drills probably won’t be heard, and the classroom would need to be notified of emergency drills in another way.

The children liked to arrange small areas for themselves from tables, moveable bulletin boards, or desks. It might be helpful to have some partitions available for them to construct their own separate areas.

It is helpful to have instrument storage that is child-accessible. This allows the students to be responsible for the care of the instruments, and also, in having the children put the room back in order, provides closure on the activity to be able to transition back to a normal school environment.

Procedural Recommendations

The children need to be secure in what is expected of them. Definite guidelines that are set from the beginning of the session as to the limits of the children's behavior prevents the supervising adult from continually intervening in behavior for disciplinary purposes. For most of the children, play in a classroom is a new experience, and they are very unsure of what to expect. The supervising adult needs to have in mind where the limits are before the play is begun.

From 20 to 30 minutes seemed to be the optimal time for musical play for groups of first, second, and third grade children. Some children could have played much longer, but others had trouble staying involved for much longer than that.

Instrument Considerations

The instruments held up well during the play treatment. It was found during the first semester of play treatment, however, that wooden maracas, guiros, and bongo drums were susceptible to breakage. The other hand drums survived just fine. The xylophones, metallophones, and glockenspiels with removable bars also survived without breakage.

Children at both schools had been trained by the regular teachers on the proper removal of the bars, and were reminded about the procedure when they obviously were interested in taking the instruments apart.

The keyboard instrument with headphones was a wonderful part of the play experience. This instrument allowed children to have a spot of relative solitude during the play time. It would be helpful to have several of these, if possible.

The students were very interested in how instruments worked. They explored the inside of the acoustic piano, and would have taken it completely apart, if they had been allowed to. It would have been helpful to have an older piano around for them to explore, as well as any other acoustic instruments that would be available for dissection.

Personal Observations on Musical Play

First, and foremost, musical play was a delightful thing to behold—in my eyes. I saw budding composition, curiosity, joy, humor, leadership, exploration of sound, experimentation, and out and out fun. However, it was chaotic, it was loud, and I believe it was the best when it verged on the edge of what many would call out of control. Indeed for free play to happen, the children must perceive that they are in control of the environment. While observing the musical play, I thought, as had many other observers of musical play whose works I have reviewed, that the educational benefits from all this uninhibited music making would be great.

However, this lack of order was viewed differently by four other educators who observed various segments of the first semester's implementation of the free play treatment (Pilot Study II). There was concern that the children, once they had

experienced such freedom in the classroom, would not tolerate “normal” music instruction. There was also concern that the freedom allowed in the music classroom, which was not the ordinary behavior allowed in the school, would cause the regular classroom teachers to have to devote more of their time to discipline. There was concern for the safety of the costly musical instruments, and concerns that the children would perceive the instruments as toys instead of as instruments capable of making fine music. There was concern that the children would learn wrong performance techniques during their play, and that fixing those bad habits would be quite time consuming. Each teacher was surprised the children did not become bored and were willing to play the whole period allotted. Three of the teachers thought the play was quite a waste of time, while the fourth saw some value in it, but did not believe the negatives outweighed any possible positive outcome. My belief, after watching a semester’s worth of musical play, was that the promise it held for musical learning far outweighed any possible problems in discipline or instrument repair.

This dichotomy between my perception of the play experience and other educators’ perception of the same experience was disturbing. It was particularly disturbing after realizing the data from this study showed no positive effect of all of this play on the quality of the students’ improvisations.

This study was birthed out of my intuition about play after watching small groups of my students come into the music classroom during their lunch periods and after school hours and ask to “play around” with the instruments. I silently observed their play. They were sometime self-conscious, but most times forgot their inhibitions and just enjoyed

their free time with instruments whose use was very much controlled during the regular music class time. They experimented with sound, figured out songs by ear, composed songs, made up dances—things I did with them in music class.

During the subsequent reading about children's play that I pursued, I became intrigued with the writings of Donald Pond about his experiences with the Pillsbury Study (Pond, 1981). His description of the behaviors of the children at play rivaled that of some of the behaviors I managed to maneuver my students into during my most wonderfully planned out lessons! I began to wonder what they could do with time and materials without my interference in their music making.

From this intuition, informal observation, and Donald Pond's spark of encouragement, a formal study was begun. I found that others had documented what children do when they play with musical materials, but no one had actually determined if that "playing around" was able to propel the children forward in their musical education. I reasoned that, if anything, improvisation was probably the aspect of musical education most likely to benefit from musical play—and set about to see if indeed that was true. The result of that quest was this study.

This research has provided information to help answer the question of whether play is beneficial for a child's growth in improvisational ability. In and of itself, free play does not appear to produce effects in total improvisational quality over the period of a semester. However, too many questions remain unanswered as to other facets of its role in the musical education of children. These questions are worthy of our efforts as researchers to provide guidance for the music educator.

APPENDIX A
CONSENT DOCUMENTS

APPENDIX A

CONSENT DOCUMENTS

Dear Parents,

The first, second, and third grades at [Name of School] Elementary School have been chosen to participate in a research project in the area of music during the spring semester of this year. Ms. Tammie Burger is heading the project as part of her doctoral studies at the University of North Texas. The project is looking at how to best help children learn to compose music. The children will be videotaped as they do their music activities at various times during the semester. The individual children's names will not be used in the final report. The information gathered from this study will be used to help teachers learn more about young children's creativity.

We are very excited to be able to participate in this project. The music activities are both fun and educational for the children. Your permission is necessary for your child to participate. Please fill out the permission slip below, indicating whether or not you would like your child to participate and return it to your child's homeroom teacher. If you have any questions about the project, please feel free to contact [names of principal, music teacher, and researcher with phone numbers given].

Sincerely,

[Principal's Name]

I *do* grant permission for my child, _____, to participate in the music research project at [Name of School] Elementary School during the Spring Semester of 1999.

I *do not* grant permission for my child, _____, to participate in the music research project at [Name of School] Elementary School during the Spring Semester of 1999.

Signed _____
(Parent or Guardian's signature)

Date _____

Institutional Review Board for Human Subjects Research Consent Letter

8- 7-00: 1:38PM;

:940+565+4277

2 / 2



University of North Texas
Sponsored Projects Administration

August 24, 1998

Tammie Burger
811 Quincy
Plainview, TX 79072

Re: Human Subjects Application No. 98-151

Dear Ms. Burger:

As permitted by federal law and regulations governing the use of human subjects in research projects (45 CFR 46), I have conducted an expedited review of your proposed project titled "The Effect of Free Play on Certain Aspects of Music Achievement." The risks inherent in this research are minimal, and the potential benefits to the subjects outweigh those risks. The submitted protocol and informed consent form is hereby approved for use of human subjects on this project.

The UNT IRB must re-review this project prior to any modifications you make in the approved project. Please contact me if you wish to make such changes or need additional information.

If you have any questions, please contact me.

Sincerely,


Walter C. Zacharias, Jr., Ed.D.
Chair, Institutional Review Board

WZ:sb

cc: IRB Members

P.O. Box 305250 • Denton, Texas 76203-5250
(940) 565-3940 • Fax (940) 565-4277 • TDD (800) 735-2989
e-mail: lane@abn.unt.edu

APPENDIX B
TESTING PROTOCOL

APPENDIX B

TESTING PROTOCOL

Instructions for Testing

Thank you so much for helping with this project—I do appreciate your time!

Below is the list of instructions—if anything is unclear, please feel free to ask questions.

1. Make sure the testing site is pre-set for the student. The instruments should be arranged to be accessible, and the video recorder should be set and focused. There should be an “X” on the floor where the student will sit to perform. The student’s number placard should be set on its own “X” so the number is visible on the video camera.
2. Consult your list and get the student to be tested from the music classroom and bring him or her to the testing site.
3. Make sure the number placard is flipped to the correct number and verify the child’s name is the one assigned to that number.
4. Ask the student to sit down on the “X”. Sit down on the floor with him or her and begin the script.

Script

Do you know the story of “Goldilocks and the Three Bears”? You are going to get to compose music that will go with the story. We will need four different pieces of music.

I have pictures here that will help us remember where the music goes (SHOW PICTURES HERE)—one for the Bears when they are happy and going out into the woods for a morning walk, one for Goldilocks when she is falling asleep in the Baby Bear’s Bed, another one for when Baby Bear finds Goldilocks has broken his chair and is angry, and the last one for Goldilocks when she is afraid and running away into the forest. Your music may be as long or as short as you like. Your music can have words, or it doesn’t have to have any words if you don’t want it to. We need four pieces of music: (POINT TO APPROPRIATE PICTURE) happy bear walking music, going to sleep music, angry bear music, and afraid, running through the forest music. I am going to leave you alone for a few minutes so you can make up your music. You may take as long as you like to make up your music. I’ll be outside in the hallway reading a book. When you are ready, let me know and we will make a performance together! I will tell the story and you can play your music at the appropriate times in the story. Let me know when you are ready.

(WAIT FOR CHILD TO SIGNAL HE OR SHE IS READY).

Now I am going to read the story, and when the time comes for your music, I’ll point to the picture so you will know it is time to play your music.

“Once upon a time there were three bears: A Papa bear, a Mama bear and a Baby bear. Mama bear had just put the porridge on the table when they decided to go for a morning walk in the beautiful sunshine (POINT TO APPROPRIATE PICTURE AND HAVE CHILD PLAY MUSIC). Very soon after they left, Goldilocks came to their house and knocked on the door. When no one answered, she went on in. She was very hungry,

so she sat down at the kitchen table to eat. She tasted Papa Bear's porridge. It was TOO hot! She tasted Mama Bear's porridge. It was too cold. She tasted Baby bear's porridge. It was just right, so she ate it all up!

“She went into the living room and looked for a place to sit down. She sat on Papa Bear's chair. It was too hard! She sat on Mama Bear's chair. It was too soft. She sat on Baby Bear's chair. It was just right! But then... it broke!

“She was sorry she broke the chair and went upstairs to take a nap. She lay on Papa Bear's bed. It was too hard. She lay on Mama Bear's bed. It was too soft. She lay on Baby Bear's bed. It was just right. She closed her eyes, and before she knew it, she drifted off to sleep (POINT TO APPROPRIATE PICTURE AND HAVE CHILD PLAY MUSIC).

“Just then, the Bears came home. Papa bear said “Someone's been eating my porridge!” Mama Bear said “Someone's been eating my porridge!” Baby Bear said: “Someone's been eating my porridge, and they ate it all up!”

“The Bears went into the living room. Papa Bear said “Someone's been sitting in my chair!” Mama Bear said “Someone's been sitting in my chair!” Baby Bear said “Someone's been sitting in my chair and they've broken it up!” Baby Bear was very angry! (POINT TO APPROPRIATE PICTURE AND HAVE CHILD PLAY MUSIC).

The Bears went upstairs. Papa Bear said “Someone's been sleeping in my bed!” Mama Bear said “Someone's been sleeping in my bed!” Baby Bear said “Someone's been sleeping in my bed, and here she is!”

Goldilocks opened her eyes and yawned after her nice nap. When she saw the bears, she was very afraid and jumped up and ran through the forest all the way home!
(POINT TO APPROPRIATE PICTURE AND HAVE CHILD PLAY MUSIC).

The end!

APPENDIX C
SUMMARY OF LESSON PLANS
AND PHYSICAL ASPECTS OF THE ROOM

APPENDIX C

SUMMARY OF LESSON PLANS AND PHYSICAL ASPECTS OF THE ROOM

Summary of Lesson Plans

The general curriculum for the first through third grades is an eclectic mix of musical literacy, performance, and composition with Orff Instruments, support of the general classroom curriculum, and movement. The following table summarizes the content covered during the regular sessions attended by both treatment and control.

Table C1

Summary of Lesson Plans

| | 1 st Grade | 2 nd Grade | 3 rd Grade |
|--------|---|--|---|
| Week 1 | Steady beat, movement, literacy preparation | Movement, same/different repertoire, 16 th notes, classroom curriculum support | Repertoire, composition, ABA form, aural recognition of rhythm |
| Week 2 | Review, rhythm, verse/refrain, PTO program | Rhythm, solfege, Orff ostinato, listening | Rhythm, solfege, Orff ostinato, listening, |

table continues

| | 1 st Grade | 2 nd Grade | 3 rd Grade |
|--------|---|--|---|
| Week 3 | PTO program, listening, dramatic movement | Improvisation and style, Mendelssohn, tempo, rhythm notation. | Mendelssohn, musical alphabet, aural timbre recognition, string family |
| Week 4 | PTO program | Orff orchestration, class curriculum support, rhythm and solfege notation | Note names, movement, rondo form, theme and variations, composition |
| Week 5 | PTO program | Vocal technique, rhythm notation, listening | Vocal technique, rhythm notation, listening |
| Week 6 | PTO program | Solfege/letter name notation, movement, listening | Listening, vocal technique, solfege, letter names, rhythm notation |
| Week 7 | PTO program | Movement, solfege, rhythm notation, listening, instrumental ostinato | Theme and variations, listening, Bach, notes and rhythm drill <u>table continues</u> |

| | 1 st Grade | 2 nd Grade | 3 rd Grade |
|---------|---|--|---|
| Week 8 | PTO program | Repertoire, movement, rhythm dictation | Listening, repertoire, note names. |
| Week 9 | Repertoire, solfege, listening | Sign language, solfege drill, timbre identification, rhythm drill | Vocal technique, repertoire |
| Week 10 | Repertoire, solfege, rhythm | Vocal technique, listening, style, solfege | Vocal technique, solfege drill, rhythm drill, classroom curriculum support |
| Week 11 | Nature sounds, composition, listening | Vocal technique, solfege, listening, music of different cultures | Listening, Solfege, repertoire |
| Week 12 | Style, creative movement, instrument accompaniment | Repertoire, round singing with instrument ostinatos, Vocal technique, solfege drill | Vocal technique, instrument ostinato, note names, solfege, repertoire. |
| Week 13 | Playing instruments, rhythm band, conducting. | Instrument aural identification, songs with movement. | Round singing with instrument ostinatos. |

table continues

| | 1 st Grade | 2 nd Grade | 3 rd Grade |
|---------|--|---|------------------------------------|
| Week 14 | Sounds around us, strong and weak beats, Orff instrumentation of songs | Repertoire, solfege and rhythm drill, movement | Vocal technique and repertoire. |

Physical Aspects of The Music Room

The classroom was carpeted, brightly lit, and large, measuring 30 feet by 30 feet. Other than the teacher's desk, file cabinets, and acoustic piano, the space was open. Storage for instruments was in two large storage rooms off the main classroom. The instruments that were used for the play treatment were the same ones utilized during regular music instruction with all groups. The instruments were placed around the room during the play treatment sessions. Students were allowed to move the instruments wherever they chose, with the exception of the acoustic and electronic pianos. The instruments available during this study are listed in Table C2.

Table C2

Inventory of Classroom Equipment

| | | |
|--------------------------|-------------------------|--------------------------|
| Acoustic Piano | Drum- 10 inch (2) | Slap stick |
| Alto Glockenspiel | Drum-8 inch (2) | Soprano Glockenspiel |
| Alto Metallophone | Drum—snare | Soprano Metallophone (2) |
| Alto Xylophone (3) | Electronic Piano | Soprano Xylophone (2) |
| Bass Xylophone | Guiro tone blocks (3) | Tambourines (3) |
| Bongo Drums | Jingle Sticks (10) | Tenor Metallophone (2) |
| Castanets (1) | Mallets (various kinds) | Tenor Xylophone (2) |
| Claves (3) | Maracas (1 pair) | Triangles (5) |
| Cymbals | Rhythm sticks (15 pair) | Wood Blocks (2) |
| Diatonic resonator bells | Sand Blocks (8 pair) | Wrist Jingles (6) |

APPENDIX D
INSTRUCTIONS FOR JUDGES AND RATING FORM

APPENDIX D

INSTRUCTIONS FOR JUDGES AND RATING FORM

Judges' Instructions

Thank you for your time in judging these children's improvisations. I do appreciate and value your input!

You should have 3 tapes (one preview tape, tape 1 and tape 2) and a notebook. The notebook contains Judges' Instructions, Adjudication Forms for tape 1 and tape 2, a reference copy of the script used to elicit the improvisational responses, and a mailing label with which to return the tapes and adjudication sheets to me.

In explanation of the project, all students were requested to improvise/compose music to accompany four specific parts of the story of Goldilocks and the Three Bears at different points during the previous semester. I have included a copy of the script used to elicit the responses to clarify the task the children were asked to complete. Your ratings of the improvisations/compositions will provide data used to evaluate teaching methods.

You will be rating all the improvisations on tapes 1 and 2 as to their quality by placing an X on the appropriate spot on a 5-point Likert scale with 1 indicating poor and 5 indicating superior. Your rating provides one number that is a holistic evaluation of the quality of all four segments. If you decide to rate a video clip between numbers, it will be interpreted at the mid-point—for example, an X placed between the numbers of 2 and 3

will be interpreted as 2.5—unless you indicate otherwise (i.e. 2.2, or 2.8, etc.). Please use the entire Likert scale to provide the best differentiation between improvisations/compositions. The subjects will not be informed of their ratings.

Please rate all video clips during one day. You may take a break between tapes, but for reliability purposes, the tapes need to be evaluated close in time to each other. The preview tape requires 15-20 minutes and the other two tapes require 2 hours each.

Specific Judging Directions

1. View the preview tape. This is to give you a general idea of the breadth of quality of the subjects' improvisations/compositions. These do not need to be rated, but are merely to orient you to the subjects' abilities.
2. View tape 1 and mark ratings during the 10-second blackout times between video clips. If you need to stop the tape and rewind for any reason, please feel free to do so. You may view each clip as many times as you desire.
3. At this point you may take a break, if desired.
4. View tape 2 and mark ratings during blackout times between video clips. If you need to stop the tape and rewind for any reason, please feel free to do so. You may view each clip as many times as you desire.
5. After finishing the evaluations, please make a copy of them for backup in case the post office misplaces your package.
6. Put all materials in the box in which they were sent, attach the mailing label, and put it in the mail to me.
7. If you have any questions, please feel free to contact me.

Judges' Rating Form

| <i><u>Number</u></i> | <i>Rating</i> | | | | |
|----------------------|----------------------|----------|----------|----------|----------|
| 359 | <u>1</u> | <u>2</u> | <u>3</u> | <u>4</u> | <u>5</u> |
| 341 | <u>1</u> | <u>2</u> | <u>3</u> | <u>4</u> | <u>5</u> |
| 364 | <u>1</u> | <u>2</u> | <u>3</u> | <u>4</u> | <u>5</u> |
| 338 | <u>1</u> | <u>2</u> | <u>3</u> | <u>4</u> | <u>5</u> |
| 623 | <u>1</u> | <u>2</u> | <u>3</u> | <u>4</u> | <u>5</u> |
| 541 | <u>1</u> | <u>2</u> | <u>3</u> | <u>4</u> | <u>5</u> |
| 252 | <u>1</u> | <u>2</u> | <u>3</u> | <u>4</u> | <u>5</u> |
| 336 | <u>1</u> | <u>2</u> | <u>3</u> | <u>4</u> | <u>5</u> |
| 542 | <u>1</u> | <u>2</u> | <u>3</u> | <u>4</u> | <u>5</u> |

APPENDIX E

DESCRIPTIONS OF JUDGES' HIGH AND LOW RATINGS

APPENDIX E

DESCRIPTIONS OF JUDGES' HIGH AND LOW RATINGS

Introduction

To provide a more complete understanding of the judges' ratings, descriptive information on examples of improvisations receiving each judge's highest and lowest scores is given. Included in the information are the ratings given by the other judges on the same improvisation and descriptive information about the music itself for each of the four sections of the improvisation requested of the children.

Highest Ratings

Judge One

The highest rating given by Judge One was a 4.0. The other two judges rated the same improvisation 4.5 and a 5.0. The Bear Walking improvisation lasted 20 seconds and was played on the xylophone using two mallets. It was in three sections. The first and last sections were similar in melody and rhythm in a 6/8 meter. The middle section consisted of several glissandi, both up and down the instrument. The piece ended with both mallets playing simultaneously. The second improvisation, Goldilocks Falling Asleep, lasted 50 seconds and was played with two mallets mainly on the glockenspiel, intermittently adding the metallophone for emphasis. A one-measure motive returned repeatedly. The end slowly died away. The third improvisation, Angry Little Bear, lasted

for 35 seconds and was in 6/8 meter. It was played with two mallets, one on the xylophone and one on the glockenspiel. There were not recurring themes and the piece ended with three loud chords. The last improvisation, Goldilocks Running Away, lasted 12 seconds. First the child used the jingle taps as a mallet to play the drum, then switched to using a clave' stick to play the drum while playing the jingle taps with his other hand. The beat was even with an accelerando and crescendo at the end.

Judge Two

The highest rating given by Judge Two was a 5.0. The other two judges rated the same improvisation 3.5 and 4.0. The Bears Walking improvisation lasted 10 seconds and was played on the xylophone and a maraca. It was in 4/4 meter and had a repeated motive. The piece had a tonal center. The second improvisation, Goldilocks Falling Asleep, lasted 5 seconds and was played with one mallet on the glockenspiel. It consisted of 2 two-measure phrases that were related melodically and rhythmically. The first one ended unresolved, and the second one ended on the tonal center. The third improvisation, Angry Little Bear, lasted 7 seconds and was played with one hand on the drum. The piece consisted of quick, steady drumbeats with accents that did not follow a metric pattern. The last improvisation, Goldilocks Running Away, lasted 7 seconds and was played by a maraca and a pair of sand blocks at the same time. It consisted of loud, even, fast beats.

Judge Three

The highest rating given by Judge Three was a 5.0. The other two judges rated the same improvisation 2.5 and 4.5. The Bears Walking improvisation lasted 55 seconds and was played on the xylophone with two mallets. The meter was in 4/4, subdivided into

triplets. The rhythm pattern repeated with a wandering melodic pattern. The improvisation began and ended with a tonal center. The second improvisation, Goldilocks Falling Asleep, lasted for 45 seconds and was played on the glockenspiel with two mallets. The tempo was slower and the dynamic level was softer. The rhythmic pattern and melodic wandering sounded similar to, but not exactly like, the Bear Walking improvisation. The third improvisation, Angry Little Bear, lasted 12 seconds and was played on the drum. The dynamics varied back and forth between loud and soft. Two mallets were used. There was a definite sense of tempo and a marked accelerando at the end. The last improvisation, Goldilocks Running Away, was played on the xylophone and lasted 25 seconds. It had rhythmic and melodic patterns similar to the first two improvisations. It was distinguished by a marked ritardando at the end.

Lowest Ratings

Judge One

The lowest rating given by Judge One was a 0.5. The other two judges rated the same improvisation 1.0 and 1.0. The child picked up a mallet but did not play an instrument for any of the improvisations. He was encouraged by the administrator, but just shook his head from side to side.

Judge Two

The lowest rating given by Judge Two was a 1.0. The other two judges also rated this improvisation 1.0. When the time came to play the first improvisation, the child stated that she didn't have music for this improvisation, but had music for the other three. On the second improvisation, the child scraped the rhythm sticks once. At the time for the

third improvisation, the child stated she didn't have music for this one either. For the fourth improvisation, the child picked up sand blocks, but then decided not to play.

Judge Three

The lowest rating given by Judge Three was a 1.0. The other two judges rated this improvisation 1.0 and 1.5. The Bears Walking improvisation lasted 10 seconds and consisted of 15 even beats played with a mallet on the drum. The second improvisation, Goldilocks Falling Asleep, consisted of 1 beat played with a mallet on the drum. The third improvisation, Angry Little Bear, lasted 13 seconds and consisted of even, quicker beats on the drum. The last improvisation, Goldilocks Running Away, lasted 6 seconds and consisted of beats on the drum that did not fall into a tempo and sped up somewhat at the end.

APPENDIX F
SUBJECTS' MEAN RATINGS

APPENDIX F

INDIVIDUAL SUBJECT'S RATINGS

Table F-1

First Grade—Control Group

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Post test |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|-------------------|
| 1 | 1.5 | 1.5 | 1.5 | 2.0 | 2.5 | 3.0 | 1.83 | 2.17 |
| 2 | 2.0 | 2.0 | 3.0 | 3.0 | 3.0 | 4.0 | 2.67 | 3.00 |
| 3 | 1.5 | 1.0 | 2.0 | 1.5 | 3.0 | 1.5 | 2.17 | 1.33 |
| 4 | 3.0 | 3.0 | 3.0 | 4.0 | 4.0 | 4.5 | 3.33 | 3.83 |
| 5 | 1.0 | 1.0 | 2.0 | 2.5 | 2.5 | 3.0 | 1.83 | 2.17 |
| 6 | 2.0 | 3.5 | 2.0 | 3.5 | 4.0 | 4.5 | 2.67 | 3.83 |
| 7 | 2.0 | 1.0 | 2.0 | 1.5 | 3.5 | 2.0 | 2.50 | 1.50 |
| 8 | 1.5 | 1.5 | 2.5 | 2.0 | 2.0 | 3.0 | 2.00 | 2.17 |
| 9 | 1.5 | 1.5 | 2.0 | 2.5 | 3.5 | 3.0 | 2.33 | 2.33 |
| 10 | 1.5 | 1.5 | 2.5 | 2.5 | 2.0 | 2.0 | 2.00 | 2.00 |
| 11 | 1.5 | 2.5 | 2.0 | 4.5 | 3.5 | 4.0 | 2.33 | 3.67 |
| 12 | 2.0 | 3.5 | 2.0 | 4.5 | 2.5 | 3.0 | 2.17 | 3.67 |
| 13 | 2.0 | 2.5 | 3.5 | 3.0 | 3.0 | 3.0 | 2.83 | 2.83 |

table continues

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Post test |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|-------------------|
| 14 | 4.0 | 4.0 | 4.5 | 4.0 | 5.0 | 5.0 | 4.50 | 4.33 |
| 15 | 1.0 | 1.5 | 1.5 | 2.5 | 2.0 | 2.0 | 1.50 | 2.00 |
| 16 | 1.0 | 1.5 | 1.5 | 2.5 | 1.5 | 3.0 | 1.83 | 2.33 |
| 17 | 2.0 | 1.5 | 2.5 | 1.5 | 4.0 | 3.5 | 2.83 | 2.17 |

Table F-2

First Grade—Treatment

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 18 | 1.0 | 1.5 | 1.5 | 3.5 | 2.5 | 4.0 | 1.67 | 3.00 |
| 19 | 1.0 | 1.0 | 2.0 | 1.5 | 2.0 | 1.0 | 1.67 | 1.17 |
| 20 | 1.5 | 2.5 | 2.0 | 2.5 | 3.0 | 3.0 | 2.17 | 2.67 |
| 21 | 2.0 | 2.0 | 3.5 | 2.5 | 3.5 | 3.0 | 3.00 | 2.50 |
| 22 | 2.0 | 2.0 | 2.0 | 1.5 | 2.0 | 3.0 | 2.00 | 2.17 |
| 23 | 2.0 | 1.5 | 2.5 | 2.0 | 3.0 | 2.5 | 2.50 | 2.00 |
| 24 | 2.0 | 1.5 | 2.5 | 2.5 | 3.5 | 2.0 | 2.67 | 2.00 |
| 25 | 1.5 | 2.5 | 2.5 | 1.5 | 3.5 | 3.5 | 2.50 | 2.50 |
| 26 | 1.5 | 1.0 | 2.5 | 2.0 | 3.0 | 3.0 | 2.33 | 2.00 |
| 27 | .5 | 1.5 | 1.0 | 3.5 | 1.0 | 3.0 | .83 | 2.67 |
| 28 | 2.0 | 2.0 | 2.5 | 2.5 | 3.0 | 3.5 | 2.50 | 2.67 |

table continues

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 29 | 2.0 | 1.5 | 2.0 | 4.5 | 3.0 | 4.0 | 2.33 | 3.33 |
| 30 | 2.5 | 2.5 | 3.0 | 3.5 | 4.0 | 3.5 | 3.17 | 3.17 |
| 31 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.00 | 1.00 |
| 32 | 2.0 | 1.5 | 4.0 | 2.5 | 3.0 | 3.5 | 3.00 | 2.50 |

Table F-3

Second Grade—Control

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 33 | 2.0 | 2.0 | 2.5 | 3.0 | 2.0 | 2.5 | 2.33 | 2.67 |
| 34 | 1.0 | 1.5 | 2.0 | 2.5 | 2.0 | 3.0 | 1.67 | 2.33 |
| 35 | 1.5 | 2.0 | 2.0 | 2.0 | 2.5 | 3.0 | 2.00 | 2.33 |
| 36 | 1.5 | 2.0 | 3.5 | 2.0 | 3.0 | 3.5 | 2.67 | 2.50 |
| 37 | 1.5 | 3.5 | 3.0 | 3.5 | 3.0 | 5.0 | 2.50 | 4.00 |
| 38 | 1.0 | 1.5 | 1.5 | 2.5 | 1.5 | 3.0 | 1.33 | 2.33 |
| 39 | 1.0 | 1.5 | 2.0 | 1.5 | 2.5 | 3.0 | 1.83 | 2.00 |
| 40 | 2.0 | 3.5 | 3.0 | 3.5 | 3.0 | 3.0 | 2.67 | 3.33 |
| 41 | 2.5 | 3.5 | 4.5 | 3.5 | 5.0 | 5.0 | 4.00 | 4.00 |
| 42 | 2.5 | 3.0 | 3.5 | 4.0 | 4.0 | 4.0 | 3.33 | 3.67 |
| 43 | 2.0 | 2.5 | 2.0 | 3.0 | 4.0 | 5.0 | 2.67 | 3.50 |
| 44 | 2.0 | 2.5 | 3.0 | 3.0 | 4.0 | 3.0 | 3.00 | 2.83 |

table continues

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 45 | 2.0 | 2.5 | 2.0 | 2.5 | 2.5 | 3.5 | 2.17 | 2.83 |
| 46 | 1.0 | 1.0 | 1.5 | 2.0 | 2.0 | 2.5 | 1.50 | 1.83 |
| 47 | 3.0 | 2.0 | 3.0 | 2.5 | 3.5 | 4.0 | 3.17 | 2.83 |
| 48 | 3.0 | 2.5 | 2.5 | 2.5 | 3.5 | 3.5 | 3.00 | 2.83 |
| 49 | 2.0 | 1.0 | 2.5 | 1.5 | 3.0 | 1.0 | 2.50 | 1.17 |
| 50 | 2.0 | 2.0 | 2.5 | 2.5 | 3.5 | 2.5 | 2.67 | 2.33 |

Table F-4

Second Grade—Treatment

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 51 | 2.0 | 2.5 | 3.0 | 2.5 | 4.0 | 3.0 | 3.00 | 2.67 |
| 52 | 1.5 | 2.0 | 2.5 | 2.5 | 4.0 | 2.5 | 2.67 | 2.33 |
| 53 | 1.0 | 2.0 | 1.5 | 3.5 | 1.5 | 3.0 | 1.33 | 2.83 |
| 54 | 2.0 | 2.0 | 2.0 | 2.0 | 3.5 | 3.5 | 2.50 | 2.33 |
| 55 | 1.0 | 1.5 | 1.5 | 2.0 | 1.0 | 3.0 | 1.17 | 2.17 |
| 56 | 1.25 | 1.5 | 2.5 | 2.0 | 2.5 | 2.0 | 2.08 | 1.83 |
| 57 | 2.0 | 2.5 | 2.0 | 2.5 | 2.5 | 3.0 | 2.17 | 2.67 |
| 58 | 2.0 | 2.0 | 3.0 | 3.0 | 3.0 | 3.0 | 2.67 | 2.67 |
| 59 | 3.0 | 3.0 | 4.5 | 3.5 | 4.5 | 5.0 | 4.00 | 3.83 |
| 60 | 2.0 | 2.5 | 2.5 | 3.5 | 3.0 | 3.0 | 2.50 | 3.00 |

table continues

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 61 | 2.5 | 2.0 | 2.5 | 2.5 | 3.0 | 3.5 | 2.67 | 2.67 |
| 62 | 1.0 | 2.0 | 2.0 | 2.0 | 2.5 | 3.0 | 1.83 | 2.33 |
| 63 | 3.0 | 2.5 | 3.5 | 3.0 | 3.0 | 1.5 | 3.17 | 2.33 |
| 64 | 4.0 | 2.5 | 3.0 | 3.5 | 3.5 | 3.5 | 2.50 | 3.17 |
| 65 | 1.5 | 2.0 | 2.0 | 2.5 | 3.5 | 3.5 | 2.33 | 2.67 |
| 66 | 3.5 | 3.0 | 5.0 | 2.5 | 4.0 | 4.5 | 4.17 | 3.33 |
| 67 | 2.0 | 1.5 | 3.5 | 2.0 | 4.0 | 2.5 | 3.17 | 2.00 |
| 68 | 1.5 | 2.5 | 1.5 | 3.0 | 3.0 | 3.5 | 2.00 | 3.00 |

Table F-5

Third Grade—Control

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 69 | 2.0 | 2.5 | 3.5 | 2.5 | 4.0 | 3.5 | 3.17 | 2.83 |
| 70 | 1.5 | 3.0 | 2.5 | 4.0 | 2.0 | 3.5 | 2.00 | 3.50 |
| 71 | 2.5 | 1.5 | 2.0 | 3.0 | 3.0 | 3.0 | 2.50 | 2.50 |
| 72 | 2.0 | 2.0 | 4.0 | 4.5 | 3.5 | 4.0 | 3.17 | 3.50 |
| 73 | 2.5 | 2.0 | 3.0 | 2.0 | 3.5 | 3.5 | 3.00 | 2.50 |
| 74 | 2.5 | 1.5 | 3.0 | 2.5 | 3.5 | 4.0 | 3.00 | 2.67 |
| 75 | 2.0 | 1.5 | 2.0 | 2.5 | 3.5 | 3.0 | 2.50 | 2.33 |
| 76 | 3.0 | 2.5 | 3.5 | 2.5 | 5.0 | 4.0 | 3.83 | 3.00 |
| 77 | 1.0 | 1.5 | 1.5 | 1.5 | 2.0 | 2.0 | 1.50 | 1.67 |

table continues

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 78 | 2.0 | 2.5 | 3.0 | 2.5 | 3.5 | 3.5 | 2.83 | 2.83 |
| 79 | 2.0 | 2.5 | 2.5 | 3.5 | 4.0 | 4.0 | 2.83 | 3.33 |
| 80 | 2.0 | 2.5 | 2.0 | 2.0 | 2.5 | 2.5 | 2.17 | 2.33 |
| 81 | 2.0 | 2.5 | 2.5 | 2.0 | 4.0 | 3.0 | 2.83 | 2.50 |
| 82 | 3.0 | 2.5 | 4.0 | 4.0 | 4.0 | 2.5 | 3.67 | 3.00 |
| 83 | 2.0 | 2.0 | 2.5 | 2.5 | 3.5 | 3.0 | 2.67 | 2.50 |
| 84 | 3.0 | 2.5 | 2.0 | 2.0 | 3.0 | 3.0 | 2.67 | 2.50 |
| 85 | 1.5 | 1.5 | 2.0 | 2.5 | 2.5 | 2.5 | 2.00 | 2.17 |
| 86 | 1.5 | 1.5 | 2.0 | 2.0 | 4.5 | 2.0 | 2.67 | 1.83 |
| 87 | 1.5 | 2.5 | 2.5 | 2.5 | 3.5 | 4.0 | 2.50 | 3.00 |
| 88 | 2.5 | 2.0 | 3.0 | 3.5 | 4.0 | 3.5 | 3.17 | 3.00 |

Table F-6

Third Grade—Treatment

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 89 | 3.0 | 2.0 | 3.5 | 3.5 | 4.0 | 4.0 | 3.50 | 3.17 |
| 90 | 3.5 | 3.0 | 2.5 | 4.0 | 4.0 | 3.5 | 3.33 | 3.50 |
| 91 | 2.5 | 2.0 | 2.5 | 2.5 | 3.5 | 3.5 | 2.83 | 2.67 |
| 92 | 2.0 | 2.0 | 1.5 | 3.0 | 2.5 | 3.0 | 2.00 | 2.67 |
| 93 | 1.0 | 1.5 | 2.0 | 2.0 | 3.0 | 4.0 | 2.00 | 2.50 |

table continues

| Subject | Judge 1 Pretest | Judge 1 Post test | Judge 2 Pretest | Judge 2 Post test | Judge 3 Pretest | Judge 3 Post test | Mean Pretest | Mean Posttest |
|---------|--------------------|----------------------|--------------------|----------------------|--------------------|----------------------|-----------------|------------------|
| 94 | 1.5 | 2.0 | 2.0 | 2.0 | 3.0 | 3.0 | 2.17 | 2.33 |
| 95 | 2.5 | 2.5 | 3.5 | 2.5 | 3.5 | 3.0 | 3.17 | 2.67 |
| 96 | .5 | 1.0 | 1.5 | 1.0 | 1.5 | 2.0 | 1.17 | 1.33 |
| 97 | 2.5 | 2.0 | 3.0 | 3.0 | 3.5 | 4.0 | 3.00 | 3.00 |
| 98 | 1.5 | 2.0 | 2.5 | 1.5 | 3.5 | 2.5 | 2.50 | 2.00 |
| 99 | 1.5 | 2.0 | 2.0 | 3.0 | 2.0 | 3.0 | 1.83 | 2.67 |
| 100 | 2.0 | 1.0 | 4.0 | 2.0 | 3.5 | 2.0 | 3.17 | 1.67 |
| 101 | 2.5 | 3.0 | 3.5 | 3.0 | 5.0 | 4.0 | 3.67 | 3.33 |
| 102 | 2.0 | 2.0 | 2.0 | 2.5 | 3.0 | 4.0 | 2.33 | 2.83 |
| 103 | 3.0 | 2.0 | 2.5 | 3.0 | 3.0 | 3.5 | 2.83 | 2.83 |
| 104 | 3.5 | 2.5 | 3.5 | 3.5 | 3.5 | 3.0 | 2.50 | 3.00 |
| 105 | 2.5 | 3.5 | 2.5 | 3.0 | 3.5 | 3.5 | 2.83 | 3.33 |
| 106 | 1.5 | 1.5 | 2.5 | 2.0 | 3.0 | 3.0 | 2.33 | 2.17 |
| 107 | 2.0 | 3.5 | 2.5 | 3.5 | 3.5 | 4.5 | 2.67 | 3.83 |
| 108 | 2.0 | 2.0 | 2.5 | 2.0 | 3.0 | 2.0 | 2.50 | 2.00 |

REFERENCE LIST

- Addison, R. (1972). A violin teaching experiment. Music in Education, 36, (355), 126-30.
- Addison, R. (1975). Children (even musical ones) make music. Music in Education, 39, (372), 60-63.
- Addison, R. (1985). In praise of the tin whistle. Music Teacher, 64, (9), 21.
- Addison, R. (1991). Music and play. British Journal of Music Education, 8, 207-217.
- Amabile, T. (1983). The social psychology of musical creativity. New York: Springer-Verlag.
- Amabile, T. & Gitomer, J. (1984). Children's artistic creativity: Effects of choice in task materials. Personality and Social Psychology Bulletin, 10, 209-215.
- Andress, B. (1986). Toward an integrated developmental theory for early childhood music education. Bulletin of the Council of Research in Music Education, 12, 10-17.
- Bangs, R. L. (1992). An application of Amabile's model of creativity to music instruction: A comparison of motivational strategies (Doctoral dissertation, University of Miami, 1992). Dissertation Abstracts International, 53, 4243.
- Brinkman, D. (1994). The effect of problem finding and creativity style on the musical compositions of high school students (Doctoral dissertation, University of Nebraska, Lincoln, 1994). Dissertation Abstracts International, 55, 2755.
- Bruner, J. S. (1972). The nature and uses of immaturity. American Psychologist, 27, 687-708.
- Bruner, J. S.; Jolly, A.; & Sylva, K. (Eds.). (1976). Play: Its role in development and evolution. New York: Penguin.
- Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.

Campbell, P. S. (1989). The significance of musical play in early childhood. General Music Today, 3, (2), (Winter), 19-20.

- Campbell, P. S. & Scott-Kassner, C. (1995). Music in childhood: From preschool through the elementary grades. New York: Schirmer Books.
- Cheyne, J. A. (1982). Object play and problem-solving: Methodological problems and conceptual promise. In D. J. Pepler & K. H. Rubin (Eds.), The play of children: Current theory and research (pp.79-96). Basel: Karger.
- Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.
- Csikszentmihalyi, M. (1979). The concept of flow. In B. Sutton-Smith (Ed.), Play and learning (pp. 257-274). New York: Gardner Press.
- Curran, J. M. (1993). The effects of pretense on the demonstration of cognitive skills (Doctoral dissertation, Arizona State University, 1993). Dissertation Abstracts International, 54, 4034.
- Dansky, J. L. & Silverman, I. W. (1973). Effects of play on associative fluency in preschool-aged children. Developmental Psychology, 9, (1), 38-43.
- Dansky, J. L. & Silverman, I. W. (1975). Play: A general facilitator of associative fluency. Developmental Psychology, 11, (1), 104.
- Ensley, S. (1987). The significance for music education of Johan Huizinga's study of the play element in culture. Canadian Journal of Research in Music Education, 29, 31-46.
- Erikson, E. (1950). Childhood and society. New York: Norton.
- Fisher, E. (1992). The impact of play on development: A meta-analysis. Play and Culture, 5, 159-181.
- Freud, S. (1942). Beyond the pleasure principle. (translated by C.J.M. Hubback) London: Hogarth Press.
- Glover, J. (1990). Understanding children's musical understanding. British Journal of Music Education, 7, (3), 257-262.
- Glover, J. & Ward, S. (Eds.). (1993). Teaching music in the primary school: A guide for primary teachers. New York: Cassell.

- Groos, K. (1898). The play of animals. New York: Appleton.
- Groos, K. (1901). The play of man. New York. Appleton.
- Hall, G. S. (1906). Youth. New York: Appleton.
- Hartmann, W. & Rollett, B. (1994). Play: Positive intervention in the elementary school curriculum. In J. Hellendoorn, R. van der Kooij, & B. Sutton-Smith (Eds.), Play and Intervention (pp. 195-202). Albany, NY: State University of New York Press.
- Hennessey, B. A. (1989). The effect of extrinsic constraints on children's creativity while using a computer. Creativity Research Journal, 2, 151-168.
- Hickey, M. (1995). Qualitative and quantitative relationships between children's creative musical thinking processes and products (Doctoral dissertation, Northwestern University, 1995). Dissertation Abstracts International, 57, 0145.
- Hinkle, D. E., Wiersma, W., & Jurs, S.G. (1988). Applied Statistics for the behavioral sciences. Boston: Houghton Mifflin.
- Huizinga, J. (1950). Homo Ludens: A study of the play element in culture. Boston: Beacon Press.
- Kartomi, M. J. (1991). Musical improvisations by children at play. The World of Music, 33, (3), 53-65.
- Li, A. K. (1978). Effects of play on novel responses in kindergarten children. Alberta Journal of Educational Research, 24, (1), 31-36.
- Lieberman, J. N. (1965). Playfulness and divergent thinking: An investigation of their relationship at the kindergarten level. Journal of Genetic Psychology, 107, 219-224.
- Lieberman, J. N. (1977). Playfulness: Its relationship to imagination and creativity. New York: Academic Press.
- Littleton, J. D. (1991). Influence of play settings on preschool children's music and play behaviors (Doctoral dissertation, University of Texas at Austin, 1991). Dissertation Abstracts International, 52, 1198.

- Marsh, K. (1995). Children's singing games: Composition in the playground? Research Studies in Music Education, 4, 2-11.
- Merrill-Mirsky, C. (1988) Eeny meeny pepsadeeny: Ethnicity and gender in children's musical play (Doctoral dissertation, University of California, Los Angeles, 1988). Dissertation Abstracts International, 49, 2445.
- Miller, L. (1983). Music in early childhood: Naturalistic observation of young children's musical behaviors (Doctoral dissertation, University of Kansas, 1983). Dissertation Abstracts International, 44, 3316.
- Mitchell, E. D. & Mason, B. S. (1934). The theory of play. New York: A.S. Barnes and Company.
- Moorhead, G., & Pond, D. (1978). Music of young children. Santa Barbara, CA: The Pillsbury Foundation for Advancement of Music Education.
- Myers, J. L. (1972). Fundamentals of experimental design, (2nd ed.). Boston: Allyn and Bacon.
- Parten, M. (1932). Social participation among preschool children. Journal of Abnormal and Social Psychology, 27, 243-369.
- Pepler, D. & Ross, H. (1981). The effects of play on convergent and divergent problem solving. Child Development, 52, 1202-1210.
- Piaget, J. (1962). Play, dreams and imitation in childhood. New York: Norton.
- Pond, D. (1981). A composer's study of young children's innate musicality. Bulletin of the Council of Research in Music Education, 68, 1-12.
- Riddell, C. (1990). Traditional singing games of elementary school children in Los Angeles (Doctoral dissertation, University of California, Los Angeles, 1990). Dissertation Abstracts International, 51, 1543.
- Rubin, K. H. , Fein, G. G., & Vandenberg, B. (1983). Play. In P. H. Mussen (Series Ed.) & E. M. Hetherington (Vol. Ed.), Handbook of child psychology: Vol 4. Socialization, personality, and social development (4th ed., pp. 693-774). New York: Wiley.

- Rubin, K.H., & Krasnor, L.R. (1980). Changes in the play behaviours of preschoolers: A short-term longitudinal investigation. Canadian Journal of Behavioural Science, *12*, 278-282.
- Sefer, J. (1995). The effects of play oriented curriculum on creativity in elementary school children. Gifted Education International, *11*, 4-17.
- Shelley, S. J. (1981). Investigating the musical capabilities of young children. Bulletin of the Council of Research in Music Education, *68*, 26-34.
- Smilansky, S. (1968). The effects of sociodramatic play on disadvantaged children. New York: Wiley.
- Simon, T. & Smith, P.K. (1985). Play and problem-solving: A paradigm questioned. Merrill-Palmer Quarterly, *31*, 265-277.
- Smith, P.K. (1995). Play, ethology, and education: A personal account. In A. D. Pellegrini (Ed.), The future of play theory (pp. 3-22). Albany, NY: State University of New York Press.
- Smith, P. K. & Dutton, S. (1979). Play and training in direct and innovative problem solving. Child Development, *50*, 830-836.
- Smith, P. K. & Simon, T. (1983). The study of play and problem solving in preschool children: Have experimenter effects been responsible for previous results? The British Journal of Developmental Psychology, *1*, 289-297.
- Smith, P. K., Simon, T. & Emberton, R. (1985). Play, problem solving and experimenter effects: A replication of Simon & Smith (1983). British Journal of Developmental Psychology, *3*, 105-107.
- Smith, P. K.; Takhvar, M.; Gore, N. & Vollstedt, R. (1985). Play in young children: Problems of definition, categorization and measurement. Early Child Development and Care, *19*, (1-2), 25-41.
- Spencer, H. (1896). Principles of psychology. (3rd ed., vol. 2, pt. 2). New York: Appleton.
- SPSS for Windows Version 8.0.0. [Computer software]. (1998). Chicago: SPSS, Inc.
- Sutton-Smith, B. (Ed.). (1979). Play and learning. New York: Gardner Press, Inc.

- Swanwick, K. (1988). Music, mind and education. New York: Routledge.
- Upitis, R. (1990). This too is music. Portsmouth, NH: Heinemann.
- Upitis, R. (1992). Can I play you my song? Portsmouth, NH: Heinemann.
- Van der Kooij, R. & Meyjes, H. (1986). Research on children's play. Prospects, 16, (1), 51-65.
- Vandenberg, B. (1981). The role of play in the development of insightful tool-using strategies. Merrill-Palmer Quarterly, 27, (2), 97-109.
- Veldhuis, H. A. (1992). Singing of preschool children in two contexts (Doctoral dissertation, The Fielding Institute, 1992). Dissertation Abstracts International, 53, 3135.
- Vygotsky, L. (1966). Play and its role in the mental development of the child. Soviet Psychology, 5, (3), 6-18.
- Vygotsky, L. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.
- Warner, B. (1991). Orff-Schulwerk: Applications for the classroom. Englewood Cliffs, NJ: Prentice Hall.
- Webster, P. & Hickey, M. (1995). Rating scales and their use in assessing children's music compositions. The Quarterly Journal of Music Teaching and Learning, 6, (4), 28-44.
- Wilson, B. (1981). Implications of the Pillsbury Foundation School of Santa Barbara in perspective. Bulletin of the Council of Research in Music Education, 68, 13-25.