

Role Play Assessments for Adults Diagnosed with Intellectual Delays:
Historical Analysis of Test Development, Construction, Application, and
Implications for Drama Therapy

Jane Maeng

A Research Paper

In

The Department

of

Creative Arts Therapies

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Arts (Drama Therapy) at
Montreal, Quebec

June 2006

© Jane Maeng, 2006



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-37776-5
Our file Notre référence
ISBN: 978-0-494-37776-5

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

ABSTRACT

Role Play Assessments for Adults Diagnosed with Intellectual Delays:
Historical Analysis of Test Development, Construction, Application, and
Implications for Drama Therapy

Jane Maeng

Drama therapy-based assessment tools are needed to evaluate client progress, create treatment plans, and support research. This paper will address developments and application of role playing assessments with adults diagnosed with intellectual delays. By using a historical analysis approach, this research aims to chronicle significant themes in psychological testing and explore implications for current drama therapy role play assessment. Developments in test construction, rating scales, testing errors, reliability and validity are analyzed within three role play assessments that are relevant to drama therapy researchers.

ACKNOWLEDGEMENTS

I would like to acknowledge my supervisor, Dr. Suzanne Lister, for her dedicated efforts in the completion of this research. I am deeply appreciative of the guidance and support she has offered. Her inspiring approach to teaching, research, and life has made this a positive experience.

I would also like to acknowledge Dr. Stephen Snow for the valued opportunity to do research in assessment with him at the Centre for the Arts in Human Development. It was a privilege to learn from a visionary in the field of drama therapy. I sincerely appreciated his expertise and assistance in guiding this research.

DEDICATIONS

This is dedicated to my loving family - Mark and Ryan. I am forever grateful for the blessings and joy they bring to my life.

My heartfelt gratitude goes to my parents, James & Winnie for their eternal support and love.

Finally, this is for my classmates, a group of nine women who have inspired and touched my life with their friendship during this journey into the Creative Arts Therapies Program.

TABLE OF CONTENTS

Heading	Page
1. Introduction	1
2. Methodology: Approaching Historical Analysis	2
a. <i>Data Collection</i>	3
b. <i>Population</i>	4
3. How Historical Analysis of Role Play Assessments is Relevant to Current Drama Therapy Practice	6
4. Interdisciplinary Research Approaches to Assessment	10
5. The Modern Testing Movement Related to Role Play Assessment and Intellectual Delays	12
6. Role Play Assessment in Psychology	15
a. <i>Structured versus Unstructured Role Play Assessments</i>	17
b. <i>Rating Scale Developments</i>	19
c. <i>Testing Errors Significant to Role Play Rating Scales</i>	23
d. <i>Reliability and Validity of Role Play Assessments</i>	25
7. Role Play Assessment #1 - Bronfenbrenner & Newcomb	27
a. <i>Advantages of Bronfenbrenner & Newcomb's Role Play Assessment</i>	29
b. <i>Challenges of Bronfenbrenner and Newcomb's Role Play Assessment</i>	32
8. Role Play Assessment #2 - Study by Kern	36
a. <i>Advantages of Kern's Role Play Assessment</i>	38
b. <i>Challenges of Kern's Role Play Assessment</i>	40

9. Role Play Assessment # 3 – David Read Johnson	43
a. <i>Strengths of DRPT – Johnson’s Role Play Assessment</i>	47
b. <i>Challenges of DRPT – Johnson’s Role Play Assessment</i>	49
10. References	54

Role Play Assessments for Adults Diagnosed with Intellectual Delays:
Historical Analysis of Test Development, Construction, Application, and
Implications for Drama Therapy

Drama therapy clinicians consider assessment an integral part of the treatment process to provide important information and evaluate client progress. As the application of drama therapy is growing relevant to a variety of populations, there is an increasing need for appropriate assessment tools. Presently, some of the most commonly known drama therapy assessments are based on role, narrative, and improvisation such as by Landy (2003), Meldrum (1994), and Johnson (1988a). Previous drama therapy assessment literature supports using qualitative and arts-based research citing that drama therapists must communicate experiences in their own language (Grainger, 1999; Landy, 1984). As drama therapy treatment involves role play, dramatic play, characterization, improvisation, and narrative, it is hypothesized that viable assessment tools would also use similar arts-based methods. Without appropriate assessment tools rooted in creative arts mediums, therapists are forced to use assessment tools from psychology or other disciplines that may fail to communicate a complete picture of personal growth occurring in clients. Psychological assessment scales are carefully developed with specific, predetermined intentions, constructs, operational definitions, and populations selected. When assessment tools are integrated into drama therapy studies without careful consideration of how aims may differ, incongruent results may be produced. This may be one of the reasons for the limited results from quantitative studies that support efficacy of drama therapy programs. The methodology employed in this research paper is historical

analysis of role play assessments used in psychology in order to support future development and research in drama therapy assessments.

This paper will address the growing need for drama therapy-based assessments. The following questions have guided this inquiry: How are role play assessments particularly relevant to the field of drama therapy? What are the specific advantages of employing role play assessment with adults diagnosed with intellectual and developmental delays? What is the meaningfulness of prior interdisciplinary research and development of role play assessments? How may historical developments in the testing movement, methodology, and theoretical frameworks influence drama therapy assessments? Furthermore, developmental issues of behavioural role play assessments created during the 1970s and 1980s such as structure, format, rating scale methods, testing errors, and psychometric properties will be examined. The last part of this research will include three role play assessments which will be analyzed on issues relevant to drama therapy applications as examples of present day concerns.

Methodology: Approaching Historical Analysis

By using a historical analysis approach, this research aims to increase familiarity with how psychological testing and role play assessment originated; thus, creating deeper understanding and meaning for researchers using current drama therapy assessment tools. Drama therapists need to understand how developments in the testing movement contributed to changes in role play assessments throughout history in order to avoid repetition of mistakes. In addition, a historical analysis of role play assessments may serve as a valuable resource of ideas for future development and interpretation. Having a

historical perspective as a researcher in drama therapy will also provide insights on the context of present challenges and advantages of role play assessments.

This research question arose while working on a project at the Centre for the Arts in Human Development with Dr. Stephen Snow. This centre, which is part of Concordia University, provides clients with developmental delays an opportunity to participate in a program consisting of several modalities of creative arts therapies. These include drama therapy, art therapy, sand play therapy, dance therapy and music therapy. Researchers from each modality selected or developed an assessment tool from their respective mediums. While assisting in the development a role play assessment for drama therapy, articles where drama therapy researchers employed the use of psychological tests in order to evaluate client progress were reviewed. Psychology is a discipline that has a prolific amount of previous research using role play assessments to evaluate adults with developmental delays. Given that drama therapy is a discipline with its own unique goals, theoretical view points, and interventions, to what extent may drama therapists rely on psychological testing methods? Is there a common ground where the two disciplines may share goals and tools?

Data Collection

In using a historical analysis approach, this research was guided by Gaye Tuchman's historical social science chapter in the *Handbook of Qualitative Research* (1994). Although I have taken courses in the history of psychology and the history of psychometrics, my method in gaining a broad background on role play assessments was partially autodidactic. Obtaining historical materials was guided by an expert in drama therapy, Dr. Stephen Snow. A comprehensive reading list was compiled by using

bibliographies of researchers who had also created role play assessments. This was fleshed out with articles from electronic databases and library catalogues searches on role play assessments from psychology, drama therapy, education and psychodrama. Some of the research topics included role play assessments, developmental delays, psychometric theory, scale development, rating scales, personality tests, improvisation and role theory. Although there have been documented cases of earlier assessments, the scope of this paper will begin in the nineteenth century as the historical developments after this period are the most significant in present drama therapy research. Historical elements will include how prior research, people and events influenced concepts of drama therapy role play assessments to present day. As a note, throughout history, developmental delays have been referred to in literature by other terms such as 'mental retardation'; however, from herein the term used will be 'people diagnosed with intellectual or developmental delays'.

Population

Clients diagnosed with intellectual and developmental delays who participate at the Centre for the Arts in Human Development display a large range in the levels of functioning and adaptability. The population is heterogeneous, with diagnoses such as Down's syndrome, Fragile X, William's syndrome, multiple sclerosis, and visual, auditory, and speech impairments. Some higher functioning individuals exhibited appropriate social behaviour, lived in semi-independent homes away from their families, and maintained employment. I was struck by certain clients who were able to communicate quite effectively. For example, people diagnosed with William's syndrome have been documented as exhibiting very talkative and outgoing characters (Udwin, Yule,

& Martin, 1987 in Hodapp & Dykens, 2003). Studies have shown that children with William's syndrome are able to tell stories using strong language abilities including sound effects and other storytelling devices (Reilly, Klima & Bellugi, 1990, in Hodapp & Dykens). For higher functioning clients, I hardly recognized that they had any impairment or delay. On the other end of the spectrum, some lower functioning clients exhibited reclusive behaviour, required prompting to participate in activities, and were very dependent on staff to get around. Some of these clients had minimal verbal communication or had perseverating speech or behaviours. The causes of intellectual and developmental delays may be due to genetic disorders, such as in the case of Down's syndrome where abnormalities occur from errors during formation of either sperm or egg (Abuelo, 1991). Fragile X is a chromosomal abnormality on the X chromosome (Abuelo). Clients diagnosed with intellectual delays may experience "difficulties attending to stimuli, learning new information and remembering old ideas" (Hale & Borkowski, 1991, p. 505).

When assessing a heterogeneous population for people diagnosed with intellectual or developmental delays, there are special needs that require consideration in the development of an assessment tool. For role playing activities, modifications need to be made to protect the necks and backs of clients with Down's syndrome. People diagnosed with severe intellectual or developmental delays have physical, ambulatory, respiratory, or heart problems which need to be accounted for while developing role play assessment tasks (Hodapp & Dykens, 2003). Test developers need to consider these special needs in constructing assessment instructions, scenarios and scales.

The criteria for diagnosing intellectual disabilities have been outlined by the American Psychological Association (APA) in the DSM-IV as well as by the American Association on Mental Retardation (AAMR). Both associations provide similar definitions for intellectual delays. The three diagnostic criteria includes an onset before age eighteen years; IQ scores below 70; and impairment in adaptive functioning in areas such as “communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skill, work, leisure, health and safety” (Hodapp & Dykens, p. 491).

How Historical Analysis of Role Play Assessments is Relevant to Current Drama Therapy Practice

Assessment and evaluation in drama therapy is a vital key for the drama therapist to reveal the client’s therapeutic needs and progress. Drama therapy assessment is usually concerned with the way in which the client uses drama and how difficulties are manifested (Jones, 1996). Assessments offer the therapist knowledge of the client and therapeutic needs (Bruscia, 1988). Jones feels that there are many questions concerning reliability and validity in drama therapy assessments. This is because drama therapy is an emergent field with little research of this nature about assessment tools (Jones). By understanding assessment tools and their historical development, drama therapists may gain more accurate information about clients by creating future research about treatment and process. This study concerns analyzing historical developments of the testing movement, theoretical frameworks, and testing aims of role play assessments in psychology to understand how they may influence future development of drama therapy assessments. There have been several articles already published summarizing existing

drama therapy assessments (see Meldrum, 1994); however, there has been no review of the substantial literature on role play assessments from psychology for the purpose of application in the field of drama therapy. Historical analysis will provide meaning to the influences of role play assessments. This topic is relevant to the field of drama therapy as drama therapists often have difficulty in providing quantitative assessment data to evaluate the effectiveness of therapeutic programs and interventions with clients when applying for grants and government funds. Meldrum writes that drama therapists are often being asked to assess and evaluate work with clients. Johnson (1988b) states, “rapid and specific assessments in each area of intervention are increasingly required for the determination of reimbursement and quality control” (p. 1). Some of the main assessment approaches used by drama therapists are interviews, projective tests and questionnaires (Bruscia; Meldrum).

Role play assessments are particularly significant to drama therapy practice for numerous reasons. Drama therapists such as Emunah (1994) and Jones (1996) propose that role play is a key process in treatment. For example, Emunah delineates that in stages of a treatment series, participants may use role play for fictional or real-life roles to “expand one’s role repertoire, foster an examination of the many aspects of one’s being, and increase one’s sense of connectedness with others” (p. 12). Through dramatization and discussion of role plays, therapeutic goals are insight-orientated but also behavioural. Clients may become more conscious of the roles they play but Emunah also proposes that in this phase, clients may learn new alternatives to behaviour (Emunah). With an expansion of role repertoire, clients may have greater self-expression, interpersonal skills, assertiveness, and coping skills (Emunah). In *Drama as Therapy*, Jones uses role playing

situations in treatment where clients prepared for life situations by using role playing as practice. He proposes that in the drama therapy process, the act of just creating a role may complete the therapeutic work (Jones). This suggests that a therapeutic component is inherently part of using role play assessment. As role play is a key factor through which therapy occurs, it would follow that role play would be a suitable medium to use in drama therapy assessment.

Role play assessments may be particularly amenable for populations such as adults with developmental delays. For example, the use of paper and pencil questionnaires may be challenging with intellectually and developmentally delayed adults due to varying literacy, comprehension, and writing skills among the population. Role play offers a unique opportunity in that it allows researchers to directly observe situational behaviour. In addition, using drama and role play has been successfully used in therapeutic processes with people diagnosed with intellectual or developmental disabilities (see Bailey, 1993; Brundenell, 1987). In fact, Bailey postulates that using drama develops and engages skills such as listening, eye contact, body awareness, physical coordination, expressiveness, problem solving, concentration and self-esteem in people with intellectual or developmental delays. In psychotherapy, conventional role play is accepted as a part of treatment which involves cognitive, emotional and behavioural components (Figge, 1982).

According to role theory, role play catalyzes development of personality and socialization. Because there are several forms of role taking, role play in assessment is applicable to all different levels of functioning participants. Mead suggests that the simplest level of role taking begins in childhood and occurs in people of all ages

(Baldwin, 1986). This developmental perspective supports role play as an appropriate medium for assessment of people with intellectual delays. During a natural progression beginning in childhood, role play acts as an influential factor in determining personality. Mead explains that in the process of taking many roles during play, a child integrates interests or characteristics he enjoys from roles to create a new sense of self (Baldwin). While learning to acquire new roles, children also begin to experience socialization with others. Through role play, children foster skills of empathy and sympathy to understand others (Baldwin). This process was studied by Pilkey, Morton, Goldman and Kleinman (1961) using psychodrama to increase empathic ability in children with intellectual or developmental delays. The authors sought to improve empathic ability so that children could overcome interpersonal difficulties. This study was an important precursor to using role play assessments in drama therapy because the relationship between empathic ability and overt ability was examined. Pilkey et al. researched play therapy, somatic therapy, and occupational therapy which were fields that also aimed to develop interpersonal skills in clients. Although the study suggested improved empathic ability, results from their rating scales may be weak considering the small sample size ($n=32$) and lack of randomized assignment. Another weakness is that the only ratings used were self-ratings by intellectually delayed children and peer ratings by other intellectually delayed children. The research findings would have been stronger if the authors had incorporated more objective raters such as teachers or parents.

One discipline with considerable historical use and development of role playing assessments is psychology. Psychology has incorporated role playing in various forms of treatment and has developed rating scale methods. A historical analysis of developments

in role play assessments and testing methodologies may offer drama therapists valuable information on future developments and interpretations of assessment research.

Interdisciplinary Approaches to Assessment

Drama therapy has similar interests and concerns as other disciplines such as psychology and education regarding assessment. Currently, researchers in these disciplines are asking similar questions such as “How may we assess personal growth as a result from treatment?”, “What are the best assessment methods to achieve our goals?” and “What are the limitations of our research due to the assessment methods we choose?” Drama therapists and psychologists often work with the same populations and create treatment programs with similar goals. Constructs prevalently considered goals of drama therapy such as interpersonal skills, social skills, assertiveness, emotional-expression, self esteem, coping skills, creativity, and quality of life have also been examined in depth by psychology. Drama therapy researchers constructing role play assessments benefit by drawing from psychometric theory and scaling methods developed in the field of psychology. Moreover, psychology has created methods to avoid testing errors that both drama therapists and psychometrists are equally susceptible to when using role play assessments or rating scales. Another commonality which drama therapists and psychologists share is that they both utilize role play in treatment and assessment. An outpour of psychological research articles on the development of role play assessments with individuals diagnosed with developmental delays occurred in the 1970s and 1980s (Bellack, 1983; Jackson, King & Heller, 1981; Senatore, Matson & Kazdin, 1982). The work of these test developers is a rich source of historical information for drama therapists who may learn from these mistakes and successes.

Although there are many similarities between drama therapy and psychology outlined above, there is an important contrast between the two fields that need to be delineated regarding theoretical frameworks and operational definitions when it comes to assessment. One difference is that the testing movement in psychology has been heavily shaped by experimentalism and behaviourism. The training programs implemented with individuals diagnosed with intellectual or developmental delays in the 1970s and 1980s tended to focus on a behavioural framework using modelling and reinforcement to teach specific skills. Consequently, the assessments created to evaluate these training programs elicited and measured these behavioural targets. In contrast, drama therapy programs have often been based on humanistic, existential or psychodynamic frameworks. Thus, assessments tended to be projective tests examining personality or designed to create treatment plans. A drama therapy role play situation may elicit desired behaviour but is usually also focused on creating developed characters, projective play, insightful interactions, life-drama connections, and therapeutic performance. Although drama therapy and psychology may be investigating similar concepts, the operational definitions of these vary from study to study. This is illustrated in the test constructs found in drama therapy and psychological research articles studying 'social skills' which often have significantly different definitions. For example, Bailey (1993), a drama therapist, describes social interaction in a therapeutic session as "sharing feelings and ideas, taking turns, being generous and kind to each other" (p. 26). On the other hand, psychology researchers such as Keller & Carlson (1974) defined social skill as the "use of generalized reinforcers in the peer group" (in Jackson, King & Heller, 1981, p. 114). Combs & Slaby (1977) defined the same concept as "ability to interact with others in a

given social context in specific ways that are socially acceptable or valued and at the same time personally beneficial, mutually beneficial or beneficial primarily to others” (in Jackson, et al., p. 114). Hersen & Bellack defined social skills as “ability to express both positive and negative feelings in the interpersonal context without suffering consequent loss of social reinforcement...[and] coordinated delivery of appropriate verbal and nonverbal responses” (1977 in Jackson, et al., p. 114). Therefore, it is important for drama therapists to consider how differences between theoretical frameworks and operational definitions will determine the usefulness of psychological assessment tools in research.

The Modern Testing Movement Related to Role Play Assessment and Intellectual Delays

Even though the first accounts of psychological testing have been credited to the Chinese empire with their exams to assign civil service 2000 years ago (Dubois, 1970), the scope of this historical analysis will focus on the major developments after the nineteenth century that have been the most influential in drama therapy assessment. As a definition, psychological tests “represent systematic applications of a few relatively simple principles in an attempt to measure personal attributes thought to be important in describing or understanding individual behaviour” (Murphy & Davidshofer, 1994, p. 1). The need to identify people with intellectual delays was fundamental in initiating the movement to develop modern psychological tests. Prior to the nineteenth century, people diagnosed with intellectual delays and mental illness commonly suffered from isolation and abuse in institutions (Anastasi & Urbina, 1997). During a period influenced by romanticism and existentialism, there was a growing humanistic interest in the treatment

and care of these populations. This necessitated the need for tools to distinguish people with developmental delays from those with mental illness as specialized social facilities were set up in Europe and the USA (Anastasi & Urbina). In the 1800s two French physicians began publishing literature on defining developmental delays and creating training programs. Esquirol created the first known description of mental retardation in 1838 and Seguin created a physiological method of training for individuals diagnosed with intellectual delays using sensory discrimination and motor control exercises which was a model implemented in institutions at that time (Anastasi & Urbina). The focus on using sensory and motor skills in assessing individuals for intellectual delays continued as experimental psychologists began to measure behaviour. Although definitions and training programs for individuals diagnosed with intellectual delays have greatly changed since Esquirol and Seguin, the approach of using science to develop testing methods has continued.

In the late nineteenth century, the rise of experimental psychology and physiology became the major influences. In 1879, Wundt founded a laboratory at Leipzig which was the training ground for future experimental psychologists (Hergenhahn, 1997). The experimental psychologists influenced present day assessments with their legacy of rigor in testing procedures, standardization, use of statistics, and focus on generalization (Hergenhahn). Researchers such as Cattell used various measures of sensory discrimination and reaction time as a test of intellectual level; however, these measures lacked validity (Anastasi & Urbina, 1997). Measuring psychometric properties such as validity and reliability continues to be an interest in modern assessment, particularly with role play assessment methods. Intelligence tests began to develop and were significant

for identification and education of children with developmental delays in schools. In 1897, Ebbinghaus created a test which sampled arithmetic, memory, and sentence completion that was significant in influencing current scales of intelligence (Anastasi & Urbina). The Binet-Simon Scale, an important development in intelligence testing, focused on reasoning, judgment, and understanding as components of intelligence, and was later refined into the Stanford-Binet Scale (Gregory, 1992). Today there are valuable intelligence tests which serve as reliable and valid tools in diagnosing developmental delays in children and adults. Lastly, World War I influenced development in psychological testing by establishing a need for rapid assessment of military personnel. During this time, Arthur S. Otis originated methods such as group testing and multiple-choice items which have continued to be used in present assessments, especially for personality testing (Anastasi & Urbina).

Role play assessments fall into the category of personality tests in psychology. Personality tests include measures of “emotional states, interpersonal relations, motivation, interests, and attitudes” (Anastasi & Urbina, 1997, p. 44). Personality assessment employs developments of standardized questionnaires, projective techniques, and rating-scale methods in test construction (Anastasi & Urbina). According to Jones (1996), projection is an important process in drama therapy treatment and can be defined as an unconscious process which places “aspects of ourselves or our feelings into other people or things” (p. 129). Personality testing allows for mediums that are conducive to drama therapy processes. For example, psychological personality tests employ dramatic projection techniques when asking a participant to do tasks such as “draw, arrange toys to

create a scene, [engage in] extemporaneous dramatic play, and interpret pictures or inkblots” (Anastasi & Urbina, p. 45).

Role playing tests are also considered situational tests in the field of personality testing. Situational tests involve engaging a participant in a situation that would be similar to their real life (Anastasi & Urbina, 1997). For example, employers may use this type of test in an interview process by asking potential employees how they would handle certain situations that may be encountered on the job. The behaviour elicited and evaluated during these tests is known as the criterion behaviour. Situation tests are attractive as they may be modified for various populations and testing goals. One of the early situational tests was by Hartshorne, May and associates for the Character Education Inquiry. This standardized test rated natural observations of children on behaviours such as honestly, self-control, and altruism in order to study development of character (Anastasi & Urbina). Popularity of situational tests continued and they were later used for selecting military personnel. During World War II, the United States Office of Strategic Services (OSS) evaluated interpersonal behaviour under stressful and frustrating conditions (Bronfenbrenner & Newcomb, 1948). These situational tests paved the way for future use of role play in assessment, especially for clients diagnosed with developmental delays.

Role Play Assessment in Psychology

Psychological literature has historically documented role playing as a technique in education, training, treatment programs, and assessment. Role play has been used in assessment for social competence of chronic psychiatric patients (Bellack, Morrison, Mueser, Wade and Sayers, 1990), social skill (Bates, 1980; Hall, Schlesinger and Dineen,

1997; Jackson, King and Heller, 1981; Merluzzi & Biever, 1987), assertiveness (Bramston, Snyder, Leah & Law, 1983), and social anxiety (Lesniack-Karpiak, Mazzocco and Ross, 2003). Role play especially became a popularized method when studying social skills, assertiveness, and interpersonal communication training programs for adults with intellectual or developmental delays. In fact, role play-based assessments have been considered the preferred method for evaluating adults with intellectual disabilities for measuring changes after training programs (Jackson, et al.). Incorporating role play as a method of assessment has largely stemmed from cognitive-behavioural researchers. This trend began in the 1970s when there was a movement to reintegrate people diagnosed with intellectual or developmental disabilities back into the community. In 1972, Wolfensberger introduced 'normalization' which aimed at providing skills to people with developmental disabilities to function closer to that of people without developmental disabilities (Jackson, et al.). What followed was an increase in research to develop training programs and role play assessments for clients diagnosed with intellectual or developmental delays. Researchers considered lack of interpersonal skills an obstacle for individuals diagnosed with developmental delays (Siperstein, 1992 in Hall, et al.) and believed training would assist integration into the community (Schalock & Harper, 1978 in Senatore, Matson & Kazdin, 1982). Behaviourally based skill training programs for people diagnosed with intellectual or developmental delays often included several components such as instructions, feedback, modelling, role playing and reinforcement (Turner, Hersen & Bellack, 1978 in Senatore, Matson & Kazdin). After creating treatment programs, an interest grew in developing role play assessments to evaluate the programs. Psychologists found role play assessments to be beneficial

because they offered a slice-of-life look into the social skills and interactions of participants. This allowed opportunities to directly observe participants in specific situations. Role play assessments are regarded as easy to administer, inexpensive, and suited to a variety of populations (Bellack, 1983).

From the 1970s and 1980s emerged several relevant topics of role play test construction. Researchers examined themes and variations of structured versus unstructured role plays; instructions; confederate behaviour; standardization; rehearsal; scoring methodology and rating scale errors. In the late 1980s and early 1990s, researchers focused on reliability and validity studies of role play assessments. These themes will be further explored and analyzed within three examples of role play assessments.

Structured versus Unstructured Role Play Assessments

An important decision for researchers constructing role play assessments in the 1970s and 1980s was whether tests should be unstructured or structured. With this decision there were implications for generalization, projection, and testing bias for either choice. Literature has produced debate about whether specific scripts for instructions and/or interactions should be used with participants. In role play assessment, confederates are often recruited to participate in improvised scenarios with clients; thus, facilitating dialogue and interaction. Employing confederates brings variables into structuring the role play scenarios. Since the nature of role play situations allow the participant endless of choices of improvised behaviour, it is challenging to create a scripted sequence of dialogue to respond to every possible situation that may arise. A consequence of having a set script is that it limits the participant's responses; thus, test

results may be negatively impacted. When developing role play situations, it is recommended that researchers allow opportunity for the criterion behaviour to be elicited (Bellack, 1983).

A benefit of having a structured role play assessment is that it allows more control over variables and bias. Merluzzi & Biever (1987) contend that standardized scripts allow for easy administration, scoring, and direct comparison. These are attractive features of assessment for researchers. Many psychological role play assessments include standardized instructions or scripts; however, some researchers have opted to use unstructured role plays to encourage improvisation. From a drama therapy perspective, an unstructured role play would allow a more natural interaction because the confederate would be able to respond directly to the participant's personal dialogue. This would allow the researcher to follow the direction the participant provides. Unstructured dialogue allows for more natural situations to be explored and an opportunity for the participant to develop their dramatic role. Although unstructured role plays encourage more spontaneity and extended interactions (Merluzzi & Biever, 1987), there is more room for the test administrator or confederate to bias interpretations or interactions by unknowingly providing cues to elicit desired behaviour. For example, the confederate may unconsciously give verbal or nonverbal direction, prompts, or cues on how the participant should respond; thus, increasing instructor demand effects. An important consideration for unstructured role play assessments is that results must be interpreted in light of possible biases and the role play situation cannot be assumed to be representative of a 'real life' situation (Bellack, 1983).

A unique drama therapy assessment tool which uses both structured and unstructured elements is Johnson's Diagnostic Role-Playing Test (1988). Johnson uses standardized instructions; however, there is no confederate or script for the role play scenario. This approach circumvents possible biases created by the confederate or test administrator. Of the various choices available for structuring, it is important to understand limitations and to interpret results with knowledge of possible implications. In addition, special training on delivering responses should be provided so that testers and confederates do not bias or influence assessment results. Therefore, structured and unstructured role play assessment procedures provide various strengths and limitations. Structured role plays allow for easy comparison, administration and limit bias. On the other hand, unstructured role plays allow for more natural improvisation, spontaneity, and allow confederates to directly respond to dialogue from the participant.

Rating Scale Developments

After deciding on whether to structure role play scenarios, scripts, and instructions, researchers must consider a method of rating to be created. Rating scales are objective, standardized descriptions which are used to "evaluate people in selection, training, psychotherapy and other interventions" (Aiken, 1996, p 20). Rating scales and checklists are frequently used by researchers to measure changes in behaviour and attitudes. These tools prove to be valuable for drama therapists evaluating clients, programs or therapeutic interventions. Ratings scales involve making evaluative judgements on a continuum representing frequencies or intensities whereas checklists are a type of rating scale with dichotomous categories (Aiken). To develop rating scales or checklists for role play assessments, there have been several proposed methods. In terms

popularized in the 1970s, molar and molecular ratings may be used to score role play interactions. Molecular measures are comprised of specific behaviours believed to be basic elements of a construct (Eisler et al., 1975 in Bellack, 1983). For example, a molecular measure may be defined as 'duration of eye contact' or 'frequency of open-ended questions asked' which would represent an element of interpersonal skill. An advantage of molecular ratings from a therapeutic view is that these allow direct feedback to clients on specific behaviours that require change (McNamara & Blumer, 1987). Molar ratings are global, overall, qualitative judgements which are presented in a single score of overall skill (Bellack, 1983). For example, a molar rating would include evaluating a role play scenario by using just one overall score for impressions of interpersonal skill. Global, molar ratings may increase a role play assessment's criterion validity; however, they create more difficulties in establishing reliability (Wessberg et al., 1979 in McNamara & Blumer, 1987).

Tests in psychology have been developed with rating scales using both molecular and molar rating systems. A criticism of molecular ratings is that they are more time consuming because they require raters to review the recorded role play several times while attending to multiple criteria. For complex constructs such as interpersonal skills, social skills, and assertiveness, molecular measures are controversial because they do not seem to support reliability when there are more than two or three measures that judges must simultaneously rate (McNamara & Blumer, 1987). Molar ratings, on the other hand, have been found to account for a significant portion of variance in role playing behaviour (Curran & Mariotto, 1981 in Bellack, Morrison, Mueser, Wade and Sayers, 1990). A problem encountered in using molar ratings is that judges will attend to particular

reference points, and easily categorized response characteristics to provide overall ratings (Trower, 1980 in Bellack 1983). Without clear references for scoring, an assessment is compromising generalizability of molar ratings because raters may be attending to different behaviours or criteria. When training raters to use molar ratings, a general explanation of a concept is often outlined; however, there is a lack of specific rating guidelines provided (Bellack, 1983). This suggests a great need for clearer guidelines and training for raters since errors often occur when judges attend to different molecular behaviours to compile global ratings (Merluzzi & Biever). In this case, the scale would be susceptible to an ambiguity error which is caused by poor descriptions and items which lack adequate information for raters (Kleinmuntz, 1982, in Aiken, 1996).

Establishing guidelines for target behaviours of assessments is important because varying outcomes in studies may be attributed to the definitions representing the concept examined. When creating role play assessment rating scales, drama therapists need to operationally define characteristics or traits by describing how they would be represented in the context of the role play scenario. There are two main construction strategies to define a construct: deductive and inductive strategies. In the deductive strategy, theoretical concepts of the behaviour studied are the main source of scale development (Burisch, 1984 in Aiken, 1996). A test developer would look at literature for theoretical relationships to the criterion being studied to create an initial list of items to be used for rating. The inductive strategy involves analyzing responses to decide the nature of the concept explored and thus, the items initially selected to make up the rating scale (Burisch, 1984 in Aiken). This strategy employs factor analysis or correlation-based procedures.

After deciding on construction methods and rating types, it is necessary to consider scale types. A technical element of role play assessment is deciding the type of rating scheme. Using simple measures such as a unidirectional rating scale to rate behaviour may be problematic for role play assessments (Bellack, 1983). For example, if an item states 'participant establishes and maintains eye contact during social interaction' and asks the rater to rank the participant where 1=little, 2=somewhat, or 3=always, it is assumed that the highest score is the optimum desired performance expected from this item. However, in evaluating a concept such as social skill, having too little or too much eye contact is not appropriate. Thus, choosing this unipolar scale of measurement would categorize an interaction where a person stares intensely or excessively for prolonged periods as indicating strong interpersonal skill. In social situations, it is not desirable to stare, or have prolonged eye contact because it may make others feel uncomfortable or be construed as aggressive. Therefore, most complex concepts such as social or interpersonal skill occur on a continuum where the optimum performance is ranked in mid range (Bellack). This is known as a bipolar rating scale where there are two extreme categories at each end and the middle represents equal amounts of each category (Aiken, 1996).

In conclusion, variables such as molecular versus molar ratings have implications for interpreting results in the context of possible bias. Thus, it is important to provide test administrators with clear guidelines on scoring. Another important factor to consider is construction strategies as a source of test items. Finally, selection of a rating scheme is crucial for communicating clear results of role play assessment data.

Testing Errors Significant to Role Play Rating Scales

Although psychological tests may occasionally be subject to errors, there are some rating errors that are significant for role play assessments in particular. In frequent cases, data is scored by multiple raters, especially in studies with large sample sizes because it may require several viewings to score the role play assessment. Role play assessment literature often refers to the significance of inter-rater agreement when there is more than one rater. Inter-rater reliability examines the scoring criteria so that it is consistently understood across raters. However, some raters may be generally inclined to constant errors where ratings are either consistently higher, lower, or more average than scores provided by other raters (Aiken, 1996). The constant error is related to a particular rater and not the scoring criteria as in problems with inter-rater reliability. To identify a constant error, researchers compare each rater's scores from the scale to the mean score of all the other raters (Aiken). In addition, an error that raters need to be cognizant of is the halo effect which represents a tendency for raters to assign favourable/ unfavourable ratings on all items because the participant demonstrates superior/ poor performance on one or two traits (Aiken). The halo effect represents a disregard for the scoring criteria in favour of certain items where the rater perceives that the participant is either more skilled or more unskilled in other areas.

A subsequent consideration for researchers to avoid rating errors is the formatting of the rating scale scoring sheet. Items may be easier to score when the rating scale is formatted in a logical fashion where categories are arranged according to theme or chronology. This assists the rater because she does not have to expend time flipping through pages to find the items while viewing the videotaped role play. However,

researchers need to examine evidence of proximity errors which occur when raters have a tendency to score similar ratings on items that are closer together on the scoring sheet. For example, on a likert scale, the rater may assign a score of 5 for a whole group of items rather than assessing the participant carefully for performance on each item. In questionnaire or survey research this would be known as a type of response set error (Aiken, 1996). One way to identify proximity errors is for the researcher to scan the scoring sheet for similar responses to a group of items. In addition, the test developer may wish to mix in reverse-scored items that will act as markers for when this error occurs.

A contrast error is the tendency to assign ratings based on the scores for the preceding participant (Aiken, 1996). Thus, rather than attending to the scoring criteria, the rater scores the present participant in relation to the performance of the previous participant. For example, if the preceding participant receives very high ratings on performance, the next participant may be assigned lower ratings in contrast because the rater does not carefully evaluate the participant according to scoring criteria (Aiken).

Lastly, the most recent performance error occurs when scores for a participant are assigned on most recent performance rather than behaviour as a whole (Aiken, 1996). To improve ratings and avoid bias, Aiken suggests providing training to raters to provide awareness on the different kinds of errors that exist. In addition, it is important to identify when these errors may be occurring and to respond by offering feedback to raters concerning their performance (Aiken).

Reliability and Validity of Role Play Assessments

After a period of rapid development of role play assessments in the 1970s and 1980s, studies evaluating reliability and validity began to emerge. In psychology the psychometric properties of role play assessments have only been investigated since the early 1980s (McNamara & Blumer, 1987). Of interest have been studies on inter-rater reliability, factorial validity, and internal consistency. In role play assessment literature, inter-rater reliability is the most published psychometric property (McNamara and Blumer). Inter-rater reliability is important in communicating the degree to which different raters agree with each other in their evaluations of behaviour in role play assessments (Rust & Golombok, 1989). To establish inter-rater reliability, raters select a sample and independently score recorded role play scenarios after familiarizing themselves with operational definitions and criterion. The scores are then correlated to provide a measure of agreement. It was the norm for assertiveness role play assessments to establish inter-rater reliabilities as high as 0.90 but coefficients in the range of 0.75 to 0.99 are acceptable (McNamara & Blumer). Inter-rater agreement has generally been found to be extremely high for checklists, rating scales, and inventories (Bates 1980; Jackson, King & Heller, 1981); however, when interpreting other types of reliability, coefficients tended to be lower than those of achievement or intelligence tests (Rust & Golombok).

Early role play assessments, such as those from the 1970s, had limited psychometric properties and relied only on face validity (Bellack, 1983). In contrast, later studies found evidence that the assessments had discriminating power between groups, were correlated with other measures, and had stable test-retest reliability (Bellack,

Morrisson, Mueser, Wade and Sayers, 1990). One of the main criticisms of role play tests is the lack of evidence to support presence of external validity (Jackson, King and Heller, 1981). Future drama therapy research may establish criterion validity by including ratings of unobtrusive observation to evaluate participants; however, collecting this data is difficult to obtain due to time constraints. In one study, Bellack, Hersen & Turner (1978) did examine external validity but found little correlation between ratings from a structured role play and behaviour from naturalistic interactions. These results suggested that role play assessments had limited value; however, the study may have found evidence of validity if researchers included a larger spectrum of behaviours in their comparisons. An example of a psychometric property study relevant to drama therapy is by Lennox (1987) where dramaturgical metaphor was used in the measurement of self-monitoring. The psychometric properties studied were face validity, internal consistency and factor analysis. This research is an excellent model for drama therapy because the assessment scale uses dramaturgy and theatre concepts to understand human interactions (Lennox). The results found evidence supporting face validity, acceptable internal consistency and one ability factor after analysis. However, a limitation to this study is that the findings lack generalizability to adults with developmental delays because the sample used was limited to undergraduate university students. Nevertheless, it is one of the few research studies relevant to drama therapy to address the psychometric properties of an assessment.

In 1987, Merluzzi & Biever continued with the psychometric research on role play assessments by publishing a validity study. Researchers compared social skills scores on a structured and an unstructured role play with a naturalistic interaction to see

whether a correlation existed. The study stated that participants found role play scenarios to be more artificial than natural interactions; however, a correlational relationship was found between scores of skill across all three types of interaction (Merluzzi & Biever). These findings support the valuable use of role play assessments. Researchers were able to rank order subjects across situations; thus arguing that role play may be substituted for naturalistic observations (Wessberg et al., 1979 in Merluzzi & Biever).

Although validity studies may provide high coefficients, researchers should assess the nature of the role play assessment before drawing conclusions. For example, if a role play assessment is unstructured, it is possible to ascertain that evidence of validity may be caused by instructor demand effects rather the test itself (Merluzzi & Biever, 1987). Thus, it is important for researchers to be aware of bias and consequent effects. Evidence of concurrent validity was supported for a role play assessment by Kern (1982) when the criterion differentiated between two extreme groups regarding anxiety and skill. This assessment will be further examined as an example, later in this research. A history of the modern testing movement related to developmental disabilities and role play assessment has been presented along with fundamental issues of structure, scoring and psychometric properties. Next, three role play assessments will be analyzed within the context of these themes and how they may be applicable to drama therapists and people diagnosed with intellectual delays.

Role Play Assessment 1 - Bronfenbrenner and Newcomb

The reason this first examined assessment by Bronfenbrenner and Newcomb was selected for analysis is because their study has been cited by both psychological and drama therapy literature. It also represents a marriage between the two fields from a very

early date in the field of role play assessment. Urie Bronfenbrenner was an established developmental psychologist (Santrock, 1996) who in 1948 published a study regarding the use of applied psychodrama in personality diagnosis. This study used six projective, improvised role play situations as an assessment for job performance. Bronfenbrenner and Newcomb conceptualized that behaviour in the assessment would reflect participants' ability to create successful interpersonal relationships (1948). Test developers included verbal and performance tasks for the assessment. The assessment is partially structured and administration consists of a brief explanation of the situation before each improvised role play situation begins (Bronfenbrenner & Newcomb). The scoring procedure involved testers writing subjective notes on each interaction. Researchers observed that using brief explanations of situations created optimal results during the improvised role plays. Bronfenbrenner and Newcomb also observed that projection of self occurred significantly more when situations differed from the circumstances of the participant's real life. The situations were based upon concepts of relationship conflicts from attachment theory regarding emotional security of participants. For example, situations involved dyadic interactions between parent and child, superior and subordinate, and opposite sex peers. After the role play situations were administered, the next step of the assessment was scoring the participants. Bronfenbrenner and Newcomb (1948) proposed the following general guide in rating participants from their notes to understand the material presented in the improvised situation:

1. How did the participant adjust to partners in the scene (parent, subordinate, member of opposite sex) and what was the characteristic mode of adjustment (dominant or submissive: outgoing or withdrawn: extropunitive or intropunitive)?

2. Did the participant show evidence of sensitivity to the feelings and actions of the partner? In creating interpersonal relationships, participants need to demonstrate that they consider needs of others.
3. Did participant provide variation of responses in the different scenes? For example, did the participant rely on one emotional response or provide an array as he embodied different roles?
4. What degree did the participant stay in each role? Testers believed that rejection of a role suggests anxiety whereas acceptance of a role suggests a source of satisfaction.
5. How did the participant regulate his/her emotions? Was there a balance or interplay between affect and intellectual control and whether emotions overrode inhibitions?
6. How did the participant end the role play situations as an aspect of character? For example, within the context of attachment theory, observers noted if there was unresolved tension, guilt, resentment, feelings of inferiority, ways of handling anxiety. For example, evidence of aggression, withdrawal, rationalization, projection, or excessive control (p. 376-378).

Advantages of Bronfenbrenner and Newcomb's Role Play Assessment

Although this test was developed almost 60 years ago, its innovation was a strong influence on shaping drama therapy role play assessments. Many of the issues raised in Bronfenbrenner and Newcomb's study are still relevant today. For example, the construct of assessing interpersonal skills through role play has been replicated by researchers in drama therapy and psychology since this study. Drama therapists have

also cited the treatment benefits of using role play as building interpersonal skills and developing relationships (Bailey, 1993; Emunah, 1994). Another relevant drama therapy concept raised in the Bronfenbrenner and Newcomb study was of spontaneity in role play situations as the authors contended the following:

[Participants] are forced into spontaneity by unforeseen and uncontrollable elements in the behaviour of his fellow participant. These elements serve to mobilize affect so that even though the situation is artificial to begin with, it soon becomes imbued with emotions that are real; that is, the subject becomes ego-involved (p. 377).

According to Moreno, a founding father of drama therapy, spontaneity is valued as an essential component of creativity and healthy living (Blatner, 2000). Researchers such as Johnson (1988), Grainger (1985), and Moreno (1946) recognize spontaneity as an important factor in role play and assessment. Bronfenbrenner and Newcomb's study further aligns itself with drama therapy methods as the unstructured tasks allow participants to have open-ended responses with limitless possibilities and creative freedom in character development, story development, and interaction with their partner. Since the instructions are minimal, the participant is able to project their inner world and feelings into the story they create. The unstructured scoring system also takes into account a wide array of responses rather than specific sets which allow testers to note more of the individual. This aspect of the assessment allows for a rich description of the participant's roles and reactions to better inform the tester of their personality. From their observations, Bronfenbrenner and Newcomb noted that "projection of self tends to occur more readily in improvisations when the subject is asked to assume a role far

removed from his actual life situations” (p. 372, 1948). The assessment uses principles in drama therapy practice by referring to the concept of what drama therapists may know as ‘paradox of projection’. The use of metaphor (the role) allows enough distance for a participant to get in touch with herself and with difficult experiences that are not accessible in other ways (Jennings, 1992). Drama therapy’s use of role play is a method that does not directly challenge a participant’s defense mechanisms. This is a key concept in drama therapy as it by-passes defense systems which are necessary for fragile identities (Jennings, Cattanach, Mitchell, Chesner, & Meldrum, 1994). In addition, through dramatic projection the client can become “emotionally and intellectually involved in encountering the problems in dramatic form” (Jones, 1996, p. 7). Projective play can bring unconscious material into awareness in a safe and playful way for greater insight.

Another advantage of this assessment is that the design permits the subject with opportunities for different reactions in varying relationships, settings, power dynamics and situations. Landy, a role theorist and drama therapist, believed that a healthy person is noted by the ability to take on many roles and playing them in everyday life (Landy, 2000). Thus, having an opportunity to perform diverse roles would be significant in role play assessment. This role play assessment has advantages over paper and pencil questionnaires concerning interpersonal relationships as the situations give observers first hand experience of how the participant is able to present herself in situations. Another advantage of this assessment is that the behaviours presented by the client are interpreted and analyzed within the context of attachment theory which is connected to interpersonal relationship skills and drama therapy. The behaviours have meaning within the theory

which allows for description and understanding of the participant. According to Bruscia (1988), Bronfenbrenner and Newcomb's assessment would be categorized as having an interpretive objective because it explains the client's behaviour in relation to the construct of attachment theory. This assessment also allows for high levels of self-disclosure where verbal communication is a metaphoric representation of experiences relating to the participant (Bruscia). Thus, the strengths of this assessment include its use of principles in drama therapy practice such as spontaneity, creative freedom, paradox of projection, and role repertoire.

This assessment would be appropriate for people diagnosed with intellectual or developmental delays for several reasons. First, Bronfenbrenner and Newcomb's assessment examines attachment and interpersonal skills of participants which are relevant topics for this population. Because the scoring system involves subjective note-taking, raters may explore issues relevant according to the individual's developmental level. This means that the scoring system does not require modifications to accommodate any special needs. The assessment is also an improvement from some of the behavioural role play assessments used with people diagnosed with intellectual delays which focused on more superficial behaviours rather than the content of the role play improvisations.

Challenges of Bronfenbrenner and Newcomb's Role Play Assessment

One challenge with Bronfenbrenner and Newcomb's assessment is that raters have their own biases, experiences, and points of view from which they are observing the participant, role, story, relationship, and situation. Thus, individual raters may see or attend to different aspects of a performance to arrive at their recorded observations. Depending on what each rater perceives, there may be conflicting results for an

individual participant from the same assessment. This is why it is important to have clear criteria to educate raters in addition to inter-rater reliability coefficients for measures of agreement. If testers do not have a viable guide then there is a greater chance for bias or unchecked influence of personal factors such as their own projections interfering with the results of the assessment.

A subsequent difficulty with Bronfenbrenner and Newcomb's assessment is that it uses very broad and general terms as guides for scoring criteria such as "balance between affect and intellectual control...do emotions tend to override inhibiting barriers?" (p. 377). The problem here is that emotions are internal phenomenon which may or may not be displayed. Another question is if a person presents affect, will an observer correctly recognize what the emotions of a participant are and the appropriate amount a participant should display? For example, a participant may be angry but a rater perceives hurt. Feelings like guilt, resentment, inferiority may be subtle differences and not accurately recognized when observing. If a participant presents with angry affect, what are the delimitations of the appropriate amount of anger to display or inhibit? One way to make this item stronger would be to interview the participants after the role play scenarios to gain their reflections on the emotions that were brought up for them while in role. In addition, "balance" is a difficult concept to discern between raters, especially when a balance is sought across many different reactions that may arise. Since there are so many possibilities, it would be almost impossible to set guidelines as to what this would look like for every possible reaction.

One problem with using Bronfenbrenner and Newcomb's assessment with people diagnosed with developmental delays is the scoring system. Although the subjective

notes used to score the scenarios allow rich descriptions of participants, more guidelines on specific behaviours are needed for special populations. For example, raters using this assessment must decide whether they are judging behaviour in relation to other people diagnosed with intellectual delays or in relation to normal populations. The standard to make comparisons against needs to be more clearly defined.

Another criterion in scoring the improvisations was whether there was varying degrees to which the participant was able to accept and stay in a particular role. Bronfenbrenner and Newcomb inferred that resistance would be manifested by breaking out of role when the participant had difficulty identifying with it; however, they did not account for the fact that not all participants will be familiar or comfortable with role playing, character development, or being on stage. In order to address this problem, a practice role play scenario should be included to acclimatize participants to the nature of role playing. Role play is a higher functioning activity that requires participants to have skill and comfort with the medium. As role play assessments are based on the assumption that situations are like “real life”, steps must be secured so that participants are given time to immerse themselves in the situations. For example, Bellack argues participants require sufficient time to imagine they are in the situations (1983). Some test developers have included time to visualize the situations or else have included practice scenes so that the participants may familiarize themselves with the structure of role play.

An additional problematic criterion for scoring the role play scenarios was whether there was variation in performance by the participant between different scenarios enacted. This may be construed as a challenge because no explanation was provided of what it meant if a person continued to use the same power relationships in every scene.

For example, because an enacted role is frequently used in improvisation, there is no evidence to suggest that the participant uses the same role dynamic dominantly in everyday relationships. From a drama therapy interpretation, other reasons a participant may be drawn to a role is because it was one he wanted to understand, wanted to rehearse, or one he modelled from another significant person in his life.

One of the key features of the assessment was the lack of structure and content. It may be argued that allowing the participant to structure his character development, and scene allows for more projective material to understand the client. It was suggested earlier that unstructured scenarios allow for a richness of interaction, spontaneity and creativity however there are other consequences. Bronfenbrenner and Newcomb believed that the most rewarding situations arose in the improvisations when minimal amounts of information regarding the situation or characters was given otherwise structuring of details of plot or role tends to “inhibit spontaneous emotional identification and increase likelihood of intellectualized or conventionalized response” (p.371). The researchers aimed to create situations where participants experienced emotion of the improvisation rather than just going through the motions (Bronfenbrenner & Newcomb). From a psychological perspective, a critical limitation of unstructured scenarios would be the lack of standardization and common criteria for basing scores because the situations were different for each participant (Bronfenbrenner & Newcomb). Because of the spontaneity and freedom offered, raters would be noting completely different content from scenes. Thus judges would have no standard method of comparing between subjects or generalizing their results.

Thus, Bronfenbrenner and Newcomb's role play assessment have unique advantages and challenges. It has been an innovative and influential tool in shaping later drama therapy assessments. For example, it incorporates drama therapy principles such as spontaneity, dramatic projection, and exploration of role repertoire. The topic of interpersonal relationships within attachment theory is also aligned to the humanistic approach of drama therapy. Challenges of this assessment include the subjectivity of the unstructured scoring system, confusing guidelines for scoring, and problematic interpretations of behaviour.

Role Play Assessment #2 - Study by Kern

This role play assessment study was chosen for analysis because it is unique in that it incorporates aspects of other role play assessment research by Bellack, Hersen, and Lamparski (1979) and Curran (1975) as well as examining unobtrusive interactions for rating. The period of time when this study was published is of importance as it occurred after a large influx of role play assessments had been developed with different ratings, types of scenarios, and goals. The dominant influence in psychology at this time was behaviourism. The purpose of Kern's (1982) study was to compare the external and concurrent validity of brief, extended, and clinical replication-type role plays for heterosocial performance. Kern's research team administered one unobtrusive interaction and three obtrusive role plays which were all rated using one proposed rating system consisting of molecular and molar measures. The first role play was a brief-response scenario developed by Bellack, et al. and involved providing standardized instructions and a confederate prompt to which the participant responded (Kern). The second scenario was an extended role play developed by Curran where participants were to enact

a first date. This involved confederates who used a standardized list of questions to prompt participants when there was a silence in order to encourage longer interactions. The third scenario was a 'replication role play', where participants are asked to re-enact previous personal experiences with the participant's real life partner (Eisler, 1976 in Kern). In this particular study, participants were asked to behave as they did in the waiting room immediately before starting the role plays with the person who was actually a confederate. All interactions were then scored for both global and specific measures of heterosocial skill and anxiety (Kern). The scoring criteria for heterosocial skill included the following specific measures:

1. Response duration – amount of time subject spoke
2. Personal attention – number of times subject used the personal pronoun "you" directed towards the confederate
3. Open-ended questions – frequency count of questions that asked for a relatively lengthy statement of opinions, feelings, or explanations
4. Closed-ended questions – frequency of questions that typically elicit a simple yes/no response or very brief response
5. Smiles – number of times subject displayed a clear upturn of the mouth (separated from previous upturns by a minimum of three seconds)
6. Loudness – average volume of the subject rated on a 1-5 anchored scale (1=very soft, 5=very loud)
7. Eye contact – duration of eye contact while speaking divided by response duration (Kern, p. 670)

In addition, raters scored a global measure of heterosocial skill which used a 7 point likert scale (extremely unskilled to extremely skilled). The global measure was guided by the specific skill measures as well other information provided by the participant (Kern).

Next, each of the interactions were rated using specific measures of observed anxiety using a modified version of the Timed Behavior Checklist (Paul, 1966 in Kern). These measures consisted of the following five categories:

1. Foot/leg movements - were repetitive movements of foot and/or legs (Foot-tapping, crossing and uncrossing of the legs)
2. Hand/arm movements – were repetitive movements of the hands and/or arms, exclusive of gestures (finger-tapping, twirling a ring)
3. Hand/arm restraint – included hand/arm positions which clearly restricted movement (sitting on hands, tight clasping of hands, clasping the edge of the seat)
4. Body movement- repetitive movements of the trunk and/or torso (fidgeting, swaying)
5. Lips – licking, biting and/or pursing of the lips (Kern, p. 671)

Kern also included a global measure of observed anxiety where raters were to attend to overall impression as well as body posture, repetitive movements and vocal qualities.

Advantages of Kern's Assessments

A unique advantage of this role play assessment study was the goal of creating a rating system that would be transferable to a variety of settings. The researcher evaluated whether long interactions were necessary or if brief interactions were sufficient to gather information about the participant. This aspect of assessment would be suitable for drama therapy clinicians who may use various types of role play interactions for evaluation. A

future implication of creating a universal rating system would be to create multiple alternative forms of an assessment. With alternative, equivalent forms of an assessment, clients would not be repeatedly exposed to the same assessment scenarios over time during a longitudinal study. Thus, learning effects would be avoided when administering the same role play scenario over time. Alternative forms would encourage spontaneity because participants would not be responding to the same scenes over and over.

Another advantages of Kern's assessment was the training offered to raters and confederates to control for bias. Lack of training for test administrators has been discussed as an issue in prior studies of role play assessment, affecting validity, reliability, and study results. As administration, interaction, and ratings may involve vulnerability to subjective factors, providing training establishes awareness and guidelines to guard against errors. In this particular study, raters were trained until criterion reliability reached .85 to .95. To control for observer drift, raters also received feedback on inter-rater reliability (Kern). In addition, confederates received extensive training and precautions were taken to control for potential changes in confederate behaviour over time (Kern). By incorporating training, Kern addresses concerns from other role play assessment researchers regarding the validity of assessments.

In prior historical debates, the pros and cons of role play assessments using molecular or molar measurements have been argued. Researchers studied which type of measure was more valid, reliable or useful. However, a clear resolution as to which is more advantageous has not been found. In this particular assessment, Kern incorporates both molecular and molar rating measures. As many assessments use one or the other, this assessment is unique in that it accounts for both ratings to compare across different

assessments. By including both types, the researcher capitalizes on the strengths of either measure.

Kern's assessment would be appropriate to use with individuals diagnosed with intellectual or developmental delays for several reasons. The role play scenarios exploring dating behaviour have relevant content for this population. As previously discussed, interpersonal relationships have widely been examined with people diagnosed with intellectual delays in research. Furthermore, the use of molecular skills will provide feedback on areas participants may need to improve.

Challenges of Kern's Assessments

One of the challenges of Kern's role play assessment is that molecular measures are time consuming. For example, Kern reported (1982) that raters were required to view each tape a minimum of five times. When sample sizes reach larger numbers then administration and rating of assessments make this assessment time inefficient. In this assessment, twelve molecular ratings were included; however, McNamara & Blumer (1987) suggest that having more than two or three affects reliability of an assessment. In addition, when using frequencies (i.e. presence or absence of molecular behaviour), rating may be difficult because scores are reliant on videotaping. If a videographer changes shots or distance, researchers may not be able to use data because they cannot accurately discern movements or perceive volumes. For example, if the camera shot is a head close-up to reveal eye contact and smiles, the camera does not allow for viewing of full body movement. Alternatively, if a shot is wide to view the entire body of the participant, more subtle behaviour such as facial movements may not be seen. Furthermore, if a different type of camera is used between participants or testing sessions,

then microphone sensitivity would also influence sound quality of the videotape which raters rely on to complete scoring. Thus, volume of sound would fluctuate between tapings. These difficulties may be relevant for proposed rating items such as smiles, loudness, and eye contact. In addition, scoring loudness would also be dependent on the raters own sense of hearing. Some raters may have a stronger sensitivity of hearing than others so scores would be difficult to compare between raters. It is necessary for the researcher to carefully develop plans and training for recording and documenting the assessments.

Another challenge of Kern's research is the fact that he has created or applied a new system of rating to pre-existing scenarios rather than using the original corresponding rating systems. This may be a problem because role play scenarios are carefully developed with specific intentions for rating. Thus, more time, analysis, and discussion on item analysis would be advantageous to support findings from their study.

Further challenges of this research include the face validity of the items used for the anxiety rating scale. For example, although the gestures and movements may have some connection to anxiety, culture and other reasons may also be causes of behaviour. In the hand/arm restraint, scorers attend to whether a participant restricts his/her movement; however, this behaviour may also be linked to cultures that often move their hands while speaking. For lip licking, the item may be measuring whether a participant is pensive, or just has dry lips. Stronger evidence for assessing validity of these items and other assessments measuring anxiety would have been desirable.

From a drama therapy perspective, the items included in this scale describe superficial behaviours and only communicate limited information regarding the many

possible facets of a client. This assessment includes measuring the amount of time a participant spoke rather than the quality or content of interactions. The behaviours examined in this assessment lack a richness of description regarding the client consequently failing to provide adequate information on how to proceed with therapeutic treatment plans.

Kern's test may have certain challenges when assessing individuals diagnosed with intellectual or developmental delays. For example, role play assessments using a behavioural framework may only provide limited information about the client because the items measure overly simplistic aspects of interpersonal skill. The scoring system is also designed for participants with language skills; thus, modifications are needed to accommodate special needs for people without verbal language.

In conclusion, one of the strengths of Kern's assessment is that it seeks to establish a universal rating system applicable to any role play situations. This has implications for future development of alternative forms of the assessment to avoid learning affects. In Kern's study, considerable training was offered to raters and confederates to avoid bias. This addressed issues of bias and testing errors prevalent in early role play assessment studies. Another advantage is that the scoring system uses both molecular and molar ratings. By incorporating both types of ratings, more information about the participants is gained. One of the challenges of Kern's assessment is that it is time consuming and relies on items difficult to score depending on the perception of the rater. In addition, further item analysis is needed and anxiety measure may lack correlation to the criterion behaviour. Lastly, this assessment is criticized for assessing superficial behaviours as a measure of social skill.

Role Play Assessment # 3 – David Read Johnson

The Diagnostic Role Playing Test by Johnson (1988) is an assessment tool which embraces the humanistic approach of drama therapy. It is unique from other assessments previously analyzed in that it views behaviour as connected to perception and the way people organize their inner world; thus, responses from assessments are suggested to predict behaviours (Blatt, 1975, in Johnson, 1988). This projective assessment aimed at psychiatric clients diagnosed with schizophrenia uses a theoretical framework based on ego psychology and object relations for personality assessment. It differs from the other psychological role play tests previously discussed because it does not directly assess interpersonal or social skills. In addition, Johnson has the background of being both a psychologist and a drama therapist.

The Diagnostic Role Playing Test (DRPT) consists of two administered role play scenarios. In the first scenario, DRPT-1, the participant is asked to enact five roles where performance is rated on presentation of an accurate and complete depiction of each social role. As the participant acts out each character, the focus is to evaluate the articulation of the stereotype of the social role. In the second scenario, DRPT -2, the participant is instructed to enact three roles that all interact with each other in the same scene. The focus of the DRPT-2 is to evaluate the participant on coherent interaction between characters. The assessment is videotaped and responses from the scenarios are transcribed. Raters are required to view the scenarios several times and assess participants using both a content and structural approach. The first approach of this assessment includes qualitative examination of content of the participant's role play to describe his or her internal world. This information may be gathered and implemented

for clinical interpretation during treatment. The second approach examines the structure of the role play by evaluating variables such as “complexity, articulation, organization and boundaries” (Johnson, p. 24). As part of clinical interpretation of the DRPT, the following diagnostic concepts of role play are analyzed:

1. Spontaneity
2. Ability to transcend reality
3. Role Repertoire
4. Organization of scenes
5. Patterns in the thematic content of scenes
6. Attitude toward enactment
7. Style of role playing (Johnson, p. 26)

In addition, developmental concepts are analyzed:

1. How does the client structure the space, tasks, and roles in the scene?
2. What media of representation (sounds, movements, images, or words) are used?
3. How complex are the characters and settings?
4. How developed are the interactions among characters?
5. What forms of affect are expressed, and to what degree? (Johnson, 1982, p. 26)

Based upon the work of Blatt and Wild (cited in Johnson), three additional scales were created based on boundary impairments thought to be discriminating factors between people diagnosed with paranoid and non-paranoid schizophrenia. The Fluid Boundary, Rigid Boundary and Complexity of Representation Scales involve the following concepts:

Fluid Boundary Scale:

1. Breaking role - momentarily coming out of role to comment on performance or engage the examiner
2. Fluidity - arbitrarily and unrepresented changes in characters, setting or objects
3. Intrusion - a foreign incongruous element is interjected into the enactment
4. Self-reference - merging within the enactment of aspects of the subject's real life and his or her role
5. Loss of distance - excessive or bizarre elaboration that suggest the subject has lost distance from the role-playing, confusing its pretend quality with reality
6. Fusion - merging or blending of aspects of different characters, objects or settings

Rigid Boundary Scale

1. Perseveration - repetition of specific characters, settings, or activities from a prior scene
2. Concrete replacement - use of a prop, wall, or floor to represent the physical presence of another person
3. Enter/leave - representation of an edge, border, limit or barrier
4. Narration - speaking in the third person about the scene
5. Telephone - use of the telephone to interact with other characters

Complexity of Representations Scale

1. Organization of scene
 - a. Developed - characters and their interaction are presented in the context of a story with a plot
 - b. Adequate - only characters and their interaction are clearly portrayed

- c. Incomplete - characters and/or plot are incompletely or vaguely portrayed
- d. Incoherent - no discernable plot, characters, or interaction

2. Integration of character/action

- a. Congruent - action is specific and congruent
- b. Nonspecific - action is not specific to role
- c. Incongruent - action is incongruous with role
- d. Fusion - only movement and gesture are portrayed

3. Interaction

- a. Active-active - both characters interact
- b. Active-reactive - one character reacts to other
- c. Active-passive - only one character is active
- d. Passive-passive - both characters are passive

4. Motivation

- a. Intentional - character has reasons for actions
- b. Reactive - character only reacts to the situation
- c. Unmotivated - reasons for actions are unclear

5. Form of character

- a. Human
- b. Quasi-human
- c. Animal
- d. Animate object
- e. Inanimate object (Johnson, 1988, p 31-33)

Strengths of the Diagnostic Role Playing Test – Johnson's Assessment

A fundamental strength of Johnson's Diagnostic Role Playing Test is the paradigm which he proposes for describing assessment results. In the context of using role within a developmental perspective, this assessment may be applicable to a wide variety of populations. Since role is a medium inherently suited to varying functioning clients, Johnson suggests that the DRPT is appropriate for people diagnosed with schizophrenia, developmental delays, children, and seniors (Johnson, 1982). Johnson describes participants along a continuum where drama therapists are able to work at each participant's respective level. This approach would be valuable for people diagnosed with intellectual delays because it describes improvisations in terms of how participants may use space, roles, media and characters in terms of developmental complexity.

Another advantage to the DRPT is that the scoring system is designed to offer an opportunity for rich description of client behaviour. Johnson's scoring system presents thoughtful descriptions of possible behaviour that may be expected during role play scenarios in addition to clinician observations. The assessment tool incorporates both qualitative descriptions as well as quantifiable measures to create a more multi-faceted depiction of participants. This is an improvement from earlier behavioural assessments which were criticized for superficiality and failure to describe the client as a whole organism. Johnson builds on previous behavioural assessment research by examining role play behaviour in addition to scenario content. Observing the subtle nuances presented in the content of scenarios such as characters, actions, symbols, or movements may comment on significant themes for the participant. This information may be relevant for future therapeutic interventions. Guidelines for clinical observations allow

drama therapists to communicate client progress with case study material. Furthermore, the developmental paradigm respects each participant's level of functioning and describes what possible progress would look like.

A subsequent benefit of Johnson's Diagnostic Role Playing Test is that the assessment is accessible and allows a trained drama therapist to be qualified to administer, interpret and analyze results. In order to do quantitative research, drama therapists must often work in collaboration with psychologists or psychometrists to use psychological tests. As opposed to using psychological tests that may require other professionals or further training, Johnson has designed this assessment to reflect the structure of a drama therapy session. Thus, the DRPT is a contribution because it user-friendly for drama therapists undertaking research.

A further advantage of the Diagnostic Role Playing Test is that Johnson has avoided problems with earlier behavioural role play assessments such as bias and superficiality. The manner in which Johnson has structured instructions is standardized and has limited bias because there is no interaction with the researcher or confederate during role play scenarios. By eliminating participant interaction with confederates, Johnson does not require a script to respond to the participants. Thus, researchers may have increased confidence that the test results are not influenced by confederates or researchers.

An additional strength of the Diagnostic Role Playing Test is that there is evidence of reliability and validity. The majority of drama therapy assessment tools lack any kind of documentation of psychometric qualities; however, Johnson suggests that this assessment is both reliable and valid. For example, Johnson provides a strong inter-rater

reliability coefficient ($p < 0.01$). There is also evidence of concurrent validity as the Fluid Boundary Scales and Rigid Boundary Scales on the DRPT correlate with the Rorschach personality test. Research has also demonstrated the DRPT has discriminating power between groups of students, and people diagnosed with paranoid and non-paranoid schizophrenia on measures of internal organization (Johnson).

Challenges of Diagnostic Role Playing Test – Johnson's Assessment

One major shortcoming of Johnson's assessment is that he claims that the Diagnostic Role Playing Test is not a complete personality assessment. He cites that the test is merely aimed at "generating initial hypotheses, ideas, directions...[for] treatment plans" (p. 27, 1988). This limits what researchers may say regarding test results for participants. Although aspects of the assessment have shown discriminating value among types of schizophrenia, results do not fulfill the goal of showing a clear picture of personality. Another limitation of the DRPT is that there is no opportunity for the drama therapist to validate interpretations of the participant's life generated from role play improvisations. This may be important because in the process of dramatic projection, the participant may use symbols, roles, dynamics, or actions that have personal meaning. One way to strengthen this assessment might be to include a structured or semi-structured interview after the scenarios to have the participant articulate the significance of the improvisation, or explain how the role play related to his reality.

When applying Johnson's model to people diagnosed with intellectual or developmental delays, is it possible for participant's to attain the highest stages of development? If the Diagnostic Role Playing Test is designed to illuminate a client's inner world, what does this mean for participants who have special needs for speech and

hearing? It is true that Johnson suggests that this test is applicable for people diagnosed with intellectual delays, but many items rely on language skills so it is difficult to implement without further amendments to the scoring system. The scoring system depends on language skills for DRPT-2 where clients are asked to create three interacting characters. If raters analyze how participants use role play to express self, is it possible to account for other methods of self-expression such as movement or sound? A clear limitation is that clients without verbal language may never achieve certain developmental levels according to this assessment. As the DRPT is a personality assessment rather than a test of interpersonal skill, a major flaw is that it may be concluded that the participant does not have a rich inner world, when in fact he just cannot verbally express it.

The Diagnostic Role Playing Test has many items that are not suitable for participants diagnosed with developmental or intellectual delays. For example, this assessment was designed for people diagnosed with Schizophrenia, so the items on the Fluid and Rigid Boundary Scales are aimed at that pathology. People diagnosed with developmental or intellectual delays do not have the same boundary issues as found in psychoses. Thus, drama therapists seeking to use this assessment with other populations may wish to further examine scale items.

The next challenge of the Diagnostic Role Playing Test is that in order to score the assessment, raters are required to transcribe role play scenario. Although analyzing verbatim transcripts is valuable for drama therapists to discover important themes or patterns, the downfall is that the assessment becomes incredibly time consuming to complete. Moreover, information gathered from transcripts may be better suited for

research comparing individual progress rather than between group progress. This limits how the DRPT may be used in certain type of drama therapy research studies.

A possible problem with the Diagnostic Role Playing Test may be the assessment's logic or validity. It is hypothesized that complex scenes, character developments, and actions reflect a participant with a well-integrated personality structure and strong internal organization. However if role playing ability is improved in clients, does this necessarily mean that a client's inner life will become enriched? For example, if professional actors proficient in performing arts were to be assessed with Johnson's DRPT, the scoring criteria would indicate they would all have a rich inner life. This would especially be true for the Complexity of Representation Scale items. Thus, the items examining content of role play scenes should have more importance than the complexity of scenes, characters or expressed affect.

Therefore, Johnson's assessment has several advantages and challenges for drama therapists. The Diagnostic Role Playing Test's developmental paradigm makes this assessment suitable for various populations. Clients are described as complex individuals because of the scoring system. Strengths of this role play test are that it is accessible, offers reliability and validity, and avoids bias. Some challenges of this assessment are that results do not reflect a complete assessment of personality. Drama therapist would require further adaptation of the DRPT in order to assess individuals diagnosed with intellectual or developmental delays because items rely on verbal language skills. This test may be time consuming because transcripts of scenarios are required to score the assessment. Issues of face validity may be problematic with items concerned with complexity of role play representations.

Conclusion and Future Research

Role play has been used as a medium in drama therapy treatment to encourage emotional expression, social skills, and expansion of role repertoire. According to role theorists, role play is suitable for various populations and is a key factor in the process of socialization and development of identity. Drama therapy and psychology researchers have constructed role playing assessments to evaluate personality, attachment styles, social skills, interpersonal skills, and assertiveness. Role play assessments offer researchers direct observation of how participants behave in situations. Historical issues of structure, rating scales, testing errors, and psychometrics from psychological literature has provided drama therapists with valuable information to interpret test results. The advantages and challenges explored in the analysis of the three assessments by Bronfenbrenner & Newcomb, Kern, and Johnson illuminate current themes relevant to drama therapy researchers using role play assessments.

Future research in the area of drama therapy role play assessments with intellectually delayed adults would benefit from further work in creating tools with an awareness of historical development. For future drama therapy role play assessments, researchers need to develop test constructs and operational definitions appropriate for the theoretical framework of the drama therapist. Rating scale items should be clearly communicated and have specific guidelines. Moreover, structuring instructions, tasks, and items need to be created with careful consideration of the specific needs of intellectually delayed adults. Further research is needed to study psychometric properties of assessments such as reliability, validity, and item analysis to increase confidence in drama therapy tools. In particular, establishing external validity continues to be an area

needing further exploration for role play assessments. Lastly, additional training to educate raters about testing errors and bias in role play assessment would be beneficial.

References

- Abuelo, D. N. (1991). Genetic disorders. In J. Matson & J. Mulick (Eds.), *Handbook of mental retardation* (pp. 97-114). New York: Pergamon Press.
- Aiken, L. R. (1996). *Rating scales and checklists: Evaluating behaviour, personality, and attitudes*. New York: John Wiley & Sons, Inc.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. New Jersey: Prentice Hall.
- Bailey, S. D. (1993). *Wings to fly: Bringing theatre arts to students with special needs*. Rockville: Woodbine House.
- Baldwin, J. D. (1986). *George Herbert Mead: A unifying theory for sociology*. London: Sage Publications.
- Bates, P. (1980). The effectiveness of interpersonal skills training on the social skill acquisition of moderately and mildly retarded adults. *Journal of Applied Behavior Analysis*, 13, 237-248.
- Bellack, A. S. (1983). Recurrent problems in the behavioral assessment of social skills. *Behaviour Research and Therapy*, 1, 29-41.
- Bellack, A. S., Hersen, M., & Lamparski, D. (1979). Role play tests for assessing social skills: Are they valid? Are they useful? *Journal of Consulting and Clinical Psychology*, 47, 335-342.
- Bellack, A. S., Hersen, M., & Turner, S. M. (1978). Role-play tests for assessing social skills: Are they valid? *Behavior Therapy*, 9(3), 448-461.
- Bellack, A. S., Morrison, R. L., Mueser, K. T., Wade, J. H., & Sayers, S. L. (1990). Role play for assessing the social competence of psychiatric patients. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 2(3), 248-255.

- Blatner, A. (2000). *Foundations of psychodrama: History, theory and practice*. New York, Springer Publishing Co.
- Bramston, P., Snyder, C. W., Leah, J. A., & Law, H. G. (1983). Assessment of assertiveness in the intellectually handicapped. *Multivariate Experimental Clinical Research*, 6(3), 143-159.
- Bronfenbrenner, U., & Newcomb, T. M. (1948). Improvisations: An application of psychodrama in personality diagnosis. *Sociatry*, 4, 367-382.
- Brudenell, P. (1987). Dramatherapy with people with a mental handicap. In S. Jennings (Ed.), *Drama therapy: Therapy & practice for teachers and clinicians* (pp. 183-208). New York: Croom Helm.
- Bruscia, K. (1988). Standards for clinical assessment in the arts therapies. *The Arts in Psychotherapy*, 15, 5-10.
- Curran, J. P. (1975). Social skills training and systematic desensitization in reducing dating anxiety. *Behaviour Research and Therapy*, 13, 65-68.
- DuBois, P. H. (1970). *A history of psychological testing*. Boston: Allyn and Bacon.
- Emunah, R. (1994). *Acting for real: Drama therapy process, technique, and performance*. Levittown: Brunner/Mazel.
- Figge, P. (1982). Dramatherapy and social anxiety: Results of the use of drama in behaviour therapy. *Dramatherapy*, 6(1), 3-17.
- Grainger, R. (1985). Using drama creatively in therapy. *Dramatherapy*, 8, 33-46.
- Grainger, R. (1999). *Researching the art therapies: A dramatherapist's perspective*. London: Jessica Kingsley Publishers.

- Gregory, R. J. (1992). *Psychological testing: history, principles, and applications*. Boston: Allyn and Bacon.
- Hale, C. A. & Borkowski, J. G. (1991). Attention, memory, and cognition. In J. Matson & J. Mulick (Eds.), *Handbook of mental retardation* (pp. 505-528). New York: Pergamon Press.
- Hall, J. A., Schlesinger, D. J. & Dineen, J. P. (1997). Social skills training in groups with developmentally disabled adults. *Research on Social Work Practice*, 7, 187-201.
- Heimberg, R. G., Montgomery, D., Madsen, C. H., Heimberg, J. S. (1977). Assertion training: A review of literature. *Behavior Therapy*, 8, 953-971.
- Hergenhahn, B. R. (1997). *An Introduction to the history of psychology*. Pacific Grove: Brooks/Cole Publishing Company.
- Hodapp, R. M. & Dykens, E. M. (2003). Mental retardation (Intellectual disabilities). In E. Mash, & R. Barkley, *Child psychopathology* (pp. 486-519). New York: Guilford Press.
- Jackson, H. J., King, N. J., & Heller, V. R. (1981). Social skills assessment and training for mentally retarded persons: A review of research. *Australian Journal of Developmental Disabilities*, 17, 113-123.
- Jennings, S. (1992). *Dramatherapy: Theory and practice 2*. London : Routledge.
- Jennings, S., Cattanach, A., Mitchell, S., Chesner, A., Meldrum, B. (1994). *The handbook of dramatherapy*. London: Routledge.
- Johnson, D. R. (1982). Developmental approaches in drama therapy. *The Arts in Psychotherapy*, 9, 183-189.

- Johnson, D. R. (1988a). The diagnostic role-playing test. *The Arts in Psychotherapy*, 15, 23-26.
- Johnson, D. R. (1988b). Introduction to the special issue on assessment in the creative arts therapies. *The Arts in Psychotherapy*, 15, 1-3.
- Jones, P. (1996). *Drama as therapy: Theatre as living*. London: Routledge.
- Kern, J. M. (1982). The comparative external and concurrent validity of three role-plays for assessing heterosocial performance. *Behavior Therapy*, 13, 666-680.
- Landy, R. J. (1984). Conceptual and methodological issues of research in drama therapy. *The Arts in Psychotherapy*, 11, 89-100.
- Landy, R. J. (2000). Role theory and the role method of drama therapy. In P. Lewis & D. Johnson (Eds.), *Current approaches in drama therapy* (pp. 50-69). Springfield: Charles C Thomas Publisher Ltd.
- Landy, R. J., Luck, B., Conner, E., & McMullian, S. (2003). Role profiles: A drama therapy assessment instrument. *The Arts in Psychotherapy*, 30 (3), 151-161.
- Lennox, R. (1987). Use of the dramaturgical metaphor in the measurement of self-monitoring. *Social Behavior and Personality*, 15(1), 111-114.
- Lesniak-Karpiak, K., Mazzocco, M. M., & Ross, J. L. (2003). Behavioral assessment of social anxiety in females with turner or Fragile X syndrome. *Journal of Autism and Developmental Disorders*, 33(1), 55-67.
- McNamara, J. R., & Blumer, C. A. (1987). Role playing: clinical assessment and therapy considerations. In J. McNamara & M. Appel (Eds.), *Critical issues, developments, and trends in professional psychology* (pp. 112-145). New York: Praeger.

- Meldrum, B. (1994). Evaluation and assessment in dramatherapy. In S. Jennings, A. Cattanach, A. Chesner, S. Mitchell, & B. Meldrum (Eds.), *The handbook of dramatherapy* (pp. 187-205). London: Routledge.
- Merluzzi, T. V., & Biever, J. (1987). Role-playing procedures for the behavioral assessment of social skill: A validity study. *Behavioral Assessment*, 9, 361-377.
- Moreno, J. L. (1946). *Psychodrama: First volume*. New York: Beacon House.
- Murphy, K. R. & Davidshofer, C. O. (1994). *Psychological testing: Principles & applications*. Englewood Cliffs: Prentice Hall.
- Pilkey, L., Goldman, M., & Kleinman, B. (1961). Psychodrama and empathic ability in the mentally retarded. *American Journal of Mental Deficiency*, 65, 595-605.
- Rust, J., & Golombok, S. (1989). *Modern psychometrics: The science of psychological assessment*. London: Routledge.
- Santrock, J. W. (1996). *Child development*. Dubuque: Brown & Benchmark.
- Senatore, V., Matson, J. L., & Kazdin, A. E. (1982). A comparison of behavioral methods to train social skills to mentally retarded adults. *Behavior Therapy*, 13, 313-324.
- Tuchman, G. (1994). Historical social science: Methodologies, methods and meanings. In N. Denzin & Y. Lincoln (Eds.), *Handbook of qualitative research* (pp. 306-332). Thousand Oaks: Sage.