Emotion Recognition and Intellectual Disability: Development of the Kinetic Emotion Recognition Assessment and Evaluation of the Emotion Specificity Hypothesis

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Clinical Psychology

at Massey University, Albany,

New Zealand.

Zara Angela Godinovich

2017

ABSTRACT

Deficits in social adaptive functioning are a defining criterion of intellectual disability (ID) (American Psychiatric Association, 2013), and a key predictor of social inclusion and subsequent quality of life (Kozma, Mansell, & Beadle-Brown, 2009). Impairment in facial emotion recognition is often cited as the component skill responsible for the social difficulties observed. This position has been formally conceptualised by the emotion specificity hypothesis (ESH; Rojahn, Rabold, & Schneider, 1995), which proposes that individuals with ID manifest a specific deficit in facial emotion recognition beyond that which can be explained by difficulties in general intellectual functioning. Despite apparent widespread acceptance, there is not yet sufficient evidence to substantiate these claims. Moore (2001) proposes that emotion perception capacities may be intact in people with ID, and that reported deficits are instead, due to emotion recognition tasks making extensive cognitive demands that disadvantage those with lesser cognitive abilities.

The aim of the present study was to clarify the nature of facial emotion recognition abilities in adults with mild ID. To this end, the Kinetic Emotion Recognition Assessment (KERA), a video-based measure of facial emotion recognition, was developed and a pilot study completed. The measure was designed to assess emotion recognition abilities, while attempting to reduce information-processing demands beyond those required to perceive the emotional content of stimuli. The new instrument was assessed for its psychometric properties in individuals with ID and neurotypical control participants. Initial findings supported the inter-rater reliability and overarching construct validity of the measure, offering strong evidence in favour of content, convergent and predictive validity. Item difficulty and discrimination analysis confirmed that the KERA included items of an appropriate level of difficulty to capture the range of emotion recognition capacities expected of individuals with mild ID.

The secondary focus of the study was to assess how subtle methodological changes in the assessment of emotion recognition ability may affect emotion recognition performance, and in turn provide insight into how we might reinterpret existing ESH literature. To this end, the KERA was also applied in an investigation of the potential moderating effects of dynamic cues and emotion intensity, in addition to the assessment of the ESH. The results offer strong evidence that individuals with ID experience relative impairment in emotion recognition abilities when compared with typically developing controls. However, it remains to be seen whether the observed difficulties are specific to emotional expression or associated with more generalised facial processing. Preliminary findings also suggest that like their typically developing peers, individuals with ID benefit from higher intensity emotional displays; while in contrast, they observe no advantage from the addition of movement cues. Finally, the overarching motivation for the reassessment and improved measurement of the ESH, was in the interests of improving real-world outcomes associated with emotion recognition capacities. Accordingly, emotion recognition data were also interpreted in the context of three measures of social functioning to explore the link between social competence and emotion recognition ability. Results indicated that emotion recognition abilities are linked to outcomes in social adaptive functioning, particularly for females.

ACKNOWLEDGEMENTS

The successful completion of this thesis is due, in no small part, to the people who have guided and shared my journey. First, I wish to thank all those who participated in this study, including the teachers, family, service providers and care staff who facilitated recruitment and participation. Thank you for your time and energy and sharing my enthusiasm for this project.

Thank you to my wonderful supervisors. To Dr Richard Fletcher, for providing invaluable theoretical input consistently from day one. Thank you for being so generous with your time and your persistence with this study. Your confidence has been a constant source of motivation throughout this project. Dr Ian de Terte, thank you for your clinical insights and your valuable feedback on written work. Your guidance and flexibility during the early stages of the project is also much appreciated. I could not have asked for two better supervisors, and I certainly hope that we have opportunity to collaborate again in the future. Others I thank are Michele Blick, who offered great support during the data collection phase of this project, and Harvey Jones of Massey University responsible for developing the computer software used in this study.

A special thank you to my family and friends, especially my partner Wilbur who has given so much so that I may complete my studies with as little stress and as much happiness as possible. I look forward to spending more time with you and many more adventures in the years ahead. Others I thank are my parents Marica and Tony, and grandparents Radica and Milan, for their ongoing care and interest in my work and general well-being. Also, Amanda for her words of wisdom and my in-laws Aubrey and Isabel for their gentle encouragement. I would also like to acknowledge my siblings and long-time friends who have patiently been there since I first decided I would like to spend another year (or two), at university. Zach, I cannot thank you enough for hours spent helping me transform a therapy room into a photo studio! Finally, thank you to my office buddies and friends made along the way, who have provided a great sense of community and comic relief. This journey would not have been the same without your friendship.

I would also like to acknowledge Massey University for their additional financial support by way of the Massey University Doctoral Scholarship and Targeted Doctoral Completion Scholarship. Thank you for your generous provision, it made a world of difference to the time and effort I was able to dedicate to this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Intellectual disability (ID), or general learning disability as it is referred to in the United Kingdom, affects between .05% and 1.55% of the population (Katherine McKenzie, Milton, Smith, & Ouellette-Kuntz, 2016). Disability manifests during the developmental years and is characterised by below-average intellectual and adaptive functioning (American Psychiatric Association, 2013). Social competence is a fundamental component of adaptive behaviour, and a key predictor of social inclusion and subsequent quality of life (Kozma et al., 2009). Deficits in facial emotion recognition are often cited as the primary component skill or causal mechanism responsible for the social adaptive difficulties experienced by individuals with ID. This position has been formally conceptualised by the emotion specificity hypothesis (ESH; Rojahn, Rabold, & Schneider, 1995), which proposes that individuals with ID of heterogeneous aetiology (excluding autism) manifest a specific deficit in facial emotion recognition beyond that which can be explained by difficulties in general intellectual functioning. The authors suggest that it may be these perceptual limitations that underlie poor social performance.

The ID literature is littered with reference to the facial emotion recognition deficits experienced by those with ID (e.g., García-Villamisar, Rojahn, Zaja, & Jodra, 2010; Zaja & Rojahn, 2008), and the existence of emotion processing deficits is accepted as the status quo to the extent that the premise is often cited as a fact rather than a hypothesis (e.g., Rojahn, Esbensen, & Hoch, 2006; Wood & Kroese, 2007). Yet despite apparent widespread acceptance, there is not yet sufficient evidence to substantiate these claims in regard to adults with ID. Moore (2000) proposes that emotion perception capacities may in fact be intact in this group, and that domain-specific emotion perception difficulties may be overstated in the literature, due to existing emotion recognition tasks making extensive cognitive demands that disadvantage people with ID. A paucity of research regarding the ecological validity of existing assessment stimuli, specifically a lack of attention to movement and colour cues and emotion intensity, complicates the issue further. Should Moore (2000) be correct in his assertion that the current findings in support of the ESH, may at least in part be an artefact of methodological limitations, then one must also re-evaluate previously observed links between facial emotion processing accuracy and social competence.

To this end, the current study sought to further clarify the nature of emotion recognition abilities of people with ID and hypothesised outcomes in social functioning. The Kinetic Emotion Recognition Assessment (KERA), a video-based measure of facial emotion, was developed and piloted. The measure was developed in the hope of improving on levels of ecological validity observed in pre-existing assessments, while attempting to reduce information-processing demands beyond those required to perceive the emotional content of stimuli. The secondary aim of the study was to assess how subtle methodological differences in the operationalisation of facial emotion may affect emotion recognition performance, and in turn provide insight into how we might reinterpret existing ESH literature. To this end, the KERA was applied in an investigation of the potential moderating effects of dynamic cues and emotion intensity, in addition to the assessment of the ESH. Emotion recognition data collected in the course of measure development and the assessment of the ESH, was then interpreted in the context of three measures of social functioning to explore the link between emotion recognition ability and social competence. Table 1.1 provides a summary of the primary and secondary aims for this study.

The first two chapters of this thesis outline the scope of the reported study. Chapter 1 presents defining criteria for ID, informed by consideration of contemporary definitions and aetiological factors and their variable influence on the study of the ESH. Chapter 2 provides the conceptual framework for the current thesis, based on discussion of two dominant emotion theories applied in experimental psychology (discrete emotion and componential appraisal theories), and how these inform the operationalisation of facial expression in emotion recognition paradigms. Chapters 3 and 4 present research relevant to the central aims of the research study. Specifically, Chapter 3 presents a select review of the literature summarising reported impairments in emotion recognition for people with ID and an examination of the research methods upon which these conclusions are based. Available emotion recognition measures are also reviewed, and a case is made for the development of a new measure in the interests of enhancing ESH research. In the context of the current thesis, determining the integrity of emotion recognition in people with ID is important in so far that it informs our theoretical understanding of social adaptive functioning in this group. To this end, Chapter 4 discusses the ESH in the wider context of social adaptive functioning, and evidence connecting emotion recognition deficits to social behaviour is reviewed.

The remainder of the thesis describes the current study. The study is introduced in Chapter 5, which provides a summary of the rationale and details the main aims. Chapter 6 and 7 describe the method and results. The results are presented in two parts: Part 1 describes the psychometric properties of the KERA; and Part 2 reports on the evaluation of the ESH and moderating factors (dynamic cues and emotion intensity), in addition to the relationship between social adaptive functioning and emotion recognition abilities. Finally, Chapter 8 presents a final discussion of the research findings and consideration of practical implications. The thesis closes with a brief discussion of the limitations of the study and recommendations for future research.

Table 1.1

Current Study Research Aims

Primary research aims	Secondary research aims
Develop and pilot a video-based measure to assess facial emotion recognition appropriate for use with people with ID.	Determine the effects of dynamic cues on facial emotion recognition performance for people with ID relative to typically developing individuals.
Re-evaluate the applicability of the emotion specificity hypothesis for individuals with ID.	Investigate the effect of emotion intensity on facial emotion recognition performance for people with ID relative to typically developing individuals.
	Explore the link between social adaptive functioning and facial emotion recognition abilities in adults with ID relative to typically developing individuals.

CHAPTER 1

DEFINING INTELLECTUAL DISABILITY

The task of defining intellectual disability (ID) is a complex one, due to the wide array of terminology and philosophical perspectives focussed on the issue. Defining criteria are developed to serve the purposes of the user, and consequently may have large implications in terms of access to services and education, legal repercussions, and personal autonomy. Accordingly, efforts are ongoing as clinicians and administrators work to develop a definition that reflects both scientific developments in the field and ethical standards related to fundamental human rights. This chapter presents three contemporary definitions of ID, a summary of known aetiology, and consideration of how aetiological factors should be accommodated in ESH research. Defining criteria for ID in the current thesis are described, informed by consideration of definitions and aetiological factors and their variable influence on the study of the Emotion Specificity Hypothesis.

Defining the diagnostic parameters of the current thesis

Contemporary systems of classification: Major criteria. A range of labels have been applied to describe people whose general intellectual functioning and adaptive coping abilities represented a departure from the usual. Within the last century diagnostic terminology has been changed approximately ten times (Harris & Greenspan, 2016), due to the societal undervaluing of those with intellectual disabilities and associated stigma leading to existing descriptors becoming pejorative (Bray & Grad, 2003). The application of early diagnostic terms such as imbecility and idiocy, developed in the 1800s, persisted throughout the first half of the twentieth century (Harris, 2006). Major reform came in 1959 when the American Association on Mental Retardation (AAMR) coined the term *mental retardation* (Heber, 1959).

The AAMR, now referred to as the American Association on Intellectual and Developmental Disabilities (AAIDD), offered the first attempt at a classification system of intellectual disability based on objective standards (e.g., test scores), and was the first classificatory scheme to be almost universally adopted (Switzky & Greenspan, 2006). The AAIDD manual, *Intellectual Disability: Definition, Classification and Systems of Support*, has since undergone 11 revisions and has subsequently informed the two major classification systems of mental disorders: The *Diagnostic and Statistical Manual of Mental Disorders (DSM)* produced by the American Psychiatric Association (APA), and the *International Classification of Diseases (ICD)* produced by the World Health Organisation (WHO). The three classification systems are broadly comparable, although significant differences remain regarding how specific criteria are operationalised and emphasis placed on personal functioning, required support provisions, and the medical underpinnings of diagnosis.

The most recent definition published by the AAIDD (Intellectual Disability: Definition, Classification and Systems of Support - 11th Edition; Schalock et al., 2010) and the APA (DSM-5; American Psychiatric Association, 2013), replace the term *mental retardation* with the more widely accepted *intellectual disability*, whilst the WHO (ICD-10; World Health Organization, 1992) continue to use *mental retardation*. Deficits in both intellectual functioning and adaptive behaviour are central to the definition of ID proposed by the three classification schemes. In addition, criterion that symptoms originate during the developmental period (i.e. prior to 18 years of age) are explicitly defined in both the AAIDD manual and DSM-5 and, while not part of the ICD-10 formal criteria, are expanded upon in supplementary ICD-10 materials (World Health Organization, 1996).

The DSM-5 and AAIDD criteria define intellectual functioning, or intelligence, according to a consensus definition that, among other things, includes the ability to learn quickly and from experience, problem solve, plan, reason, think abstractly, and understand complex ideas (Gottfredson, 1997). The ICD-10 criteria for mental retardation does not offer a specific definition of intellectual functioning; however, supplementary materials make reference to cognitive functions including memory, language, perception, and the ability to learn (World Health Organization, 1992). There is acceptance across the three classification systems that intellectual capacity should be inferred from the intelligence quotient (IQ), based on standardised individually administered tests. Significantly sub-average general intellectual functioning approximates total scores two standard deviations below the population mean. For a test with a mean of 100 and standard deviation of 15, total scores below 70 meet criteria (Wechsler, 2008).

Despite sharing a common definition for what constitutes significant limitations in intellectual function, the limits of 'normal' intelligence are interpreted slightly differently across the three classification schemes. The AAIDD and the DSM-5 both advocate for an approximate cut-off score of 70 with strong emphasis that allowances be made for standard error of measurement and instrument strengths and limitations. Accordingly, both systems accept an upper limit of approximately 75. In contrast the ICD-10 sets the upper limit at 69. While a difference of five IQ points appears slight, application of the more liberal AAIDD and DSM-5 criteria compared with the more conservative ICD-10, may lead to a twofold increase in individuals who meet criteria for limitations in intellectual functioning and a subsequent increase in the prevalence of ID (National Research Council (US) Committee on Disability Determination for Mental Retardation, 2002).

The second essential feature of intellectual disability is impairment in everyday adaptive functioning. Adaptive functioning refers to how an individual is able to meet cultural standards of personal independence and social responsibility, when evaluated against peers of similar age and sociocultural background (Tassé, 2009). Comparable to definitions of intellectual functioning, the DSM-5 and AAIDD have a shared conceptualisation of adaptive functioning which describes a collection of social, conceptual and practical skills typically learned during the developmental period. However, here too lie differences in how the construct is operationalised. The AAIDD requires that significant delay should be observed in the development of one or a combination of social, conceptual, and practical skills, and confirmed by scores two standard deviations below the mean on standardised ratings of adaptive behaviour. In contrast, the DSM-

5 emphasises the wider impact of identified limitations, and holds that adaptive deficits in any of the three domains should limit functioning in one more activities of daily life across multiple environments. The ICD-10 does not offer specific domains of functioning; rather it points to diminished social competence as the central feature.

Subclassification and severity specifiers. Intellectual disability is a highly heterogeneous concept, and individuals belonging to this group display huge variation in terms of their intellectual and adaptive functioning (American Psychiatric Association, 2013). At the more able end of the spectrum exist individuals whose behavioural, social, and emotional difficulties are more closely aligned with people of average intelligence than to individuals at the middle or lower bounds of the spectrum (World Health Organization, 1996). To accommodate such difference, sub-classification of ID is useful and necessary for research purposes.

The AAIDD 11th Edition manual describes a multi-dimensional classification system applying a supports-based paradigm. Cases are classified based on the level of supports required (intermittent, limited, extensive and pervasive) across five domains: intellectual disabilities, health, adaptive behaviour, participation and social context. While the AAIDD approach to classification is perhaps the most progressive available, critics argue that it is important to specifically emphasise intellectual or adaptive functioning capacities, and by focusing simply on required supports one discounts the differences in adaptive functioning and intellectual characteristics that dictate day-to-day functioning (MacMillan, Gresham, & Siperstein, 1993). It is arguably these characteristics that allow the researcher or clinician to not only predict domains requiring support, but also develop reasonable expectations regarding the limits of what the individual may achieve given adequate support. Further, there is considerable overlap between AAIDD level of support requirements and traditional IQ score groupings (Buntinx et al., 2008), indicating that support requirements may not be considered a particularly accurate proxy for intellectual functioning. Accordingly, the prevailing approach used by psychological researchers in further defining ID diagnosis is by the severity of observed impairment rather than support requirements.

The ICD-10 and DSM-5 apply an impairment based system of classification defined by the descriptors mild, moderate, severe and profound. The ICD-10 present IQ levels as 'guides' to categorisation (mild 50-69, moderate 35-49, severe 20-34 and profound being below 20), while sub-classification within the DSM-5 system is based on limitations in adaptive functioning within the social, practical and conceptual domains. Adaptive functioning is a broad construct used to describe the degree to which an individual meets common standards of personal independence and social responsibility (American Psychiatric Association, 2013). Only in the latest revision of the DSM did the APA shift its emphasis from classification based on IQ scores to adaptive functioning. While adaptive functioning is arguably more difficult to classify into discrete groupings, the shortcomings of an IQ-based system are well documented.

The most significant limitation of classification by IQ is that the first diagnostic criterion for ID, limitations in intellectual functioning, does not map neatly onto an IQ score (Harris & Greenspan, 2016). Specifically, gold standard intelligence tests such as the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS–IV; Wechsler, 2008) and Stanford-Binet – Fifth Edition (Roid, 2003), by no means tap all of the constituent elements of intelligence, particularly subdomains related to executive function. Further, while intellectual functioning and adaptive functioning are interrelated, the relationship is by no means perfect (National Research Council (US) Committee on Disability Determination for Mental Retardation, 2002), and the correlation between adaptive functioning and intelligence becomes less pronounced in more mild cases (National Research Council (US) Committee on Disability Determination for Mental Retardation, 2002). Accordingly, the severity of an individual's presentation or needs can only be accurately established based on adaptive functioning. It is the assessment of adaptive functioning that can determine the degree of mastery of practical tasks or reasoning in real life situations (American Psychiatric Association, 2013).

Diagnostic parameters for the current thesis. Intellectual disability as a diagnostic category poses an amorphous concept informed both by social convention and scientific discovery. It must be recognised that ID research is being conducted on a changing landscape, not only due to ongoing aetiological discoveries but also regarding how key criteria are operationalised amongst available classification systems. Therefore, defining the diagnostic parameters of enquiry must be made explicit. Among the three classification systems presented, both the DSM-5 and AAIDD overcome the described limitations associated with subcategorisation according to IQ scores. However, the limiting factor of the AAIDD is the complexity of the multidimensional approach to subclassification. Describing participants across such a wide number of dimensions can impede research by making it difficult to isolate homogenous groups of people for study (Weis, 2013). Further, the measurement of required support provisions integral to the AAIDD system arguably are one step removed from the adaptive behaviours and intellectual capacities that inform them. These capacities are central to the current investigation. Accordingly, the current study is defined by the definition of intellectual disability assigned by the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders – Fifth edition (DSM-5; American Psychiatric Association, 2013) (Appendix A). Methodology and interpretation of findings will be considered exclusively within the DSM-5 system; however, for completeness all relevant ESH literature, irrespective of defining criteria, will be subject to review.

Defining the aetiological parameters of the current thesis

The organic versus cultural-familial (nonspecific) division. Intellectual disability is an outcome state which may result from a wide range of causes, many of which remain unknown to the affected individual and scientific community alike. Widely accepted as a neurodevelopmental disorder, known causes are due to deficient or atypical brain development associated with one or a combination of prenatal (e.g., single-gene conditions, chromosomal abnormalities or maternal alcohol consumption), perinatal (e.g., labour and delivery-related events leading to neonatal encephalopathy), and post-natal (e.g., lack of environmental stimulation, malnutrition and traumatic brain injury) risk factors (Carr, 2016). Despite there being over 740 known causes of ID (Weis, 2013), very few demonstrate a one-to-one relationship with intellectual disability and approximately 60% percent of cases of ID do not have a known etiology (Ellison, Rosenfeld, & Shaffer, 2013).

Long present in the field of intellectual disability research, a 'two-group' distinction has been applied to describe the neurodevelopmental underpinnings of ID. Historically, individuals with ID were separated into two groups, those with an identifiable *organic* cause for their impairments, such as an acquired or genetic impairment, and those without, referred to as the *familial* group (Zigler, 1969). Within this system, individuals with 'organic' disability typically experienced more severe impairment, with IQ scores below 50, in addition to medical complications and physical features suggestive of underlying neurological complications. These individuals typically came from all socioeconomic backgrounds and had parents and siblings of normal intellectual functioning (Iarocci & Petrill, 2012). Children in the familial group experienced more mild presentations, with IQ scores between 50 and 70, normal physical appearance and no additional health or medical conditions. They were however, more likely to come from low-income families with a higher incidence of low intellectual functioning in biological relatives (Weis, 2013). Today, in light of recent gains regarding the nature-nurture interplay in the development of ID, this group is more commonly referred to as the 'cultural-familial' group. Intellectual and behavioural difficulties are now understood to be the culmination of environmental deprivation (e.g., low levels of cognitive stimulation) and a genetic diathesis toward low intelligence (Iarocci & Petrill, 2012).

The organic versus cultural-familial division and associated inferences regarding impairment severity hold true in many cases, and offer a broad-brush approach with which to describe the underlying aetiology of ID. However, more recent research highlights significant heterogeneity and overlap regarding the underlying pathology and the severity of difficulties experienced across both groups. In more than 50% of cases more than one causal factor is implicated (Ainsworth & Baker, 2004), and often both psychosocial and medical risk factors are identified (Carr, 2016). Further, within the organic group there are numerous unrelated causes associated with ID that cannot be assumed to result in equivalent impairment both within and across aetiological groups (Burack, Hodapp, Iarocci, & Zigler, 2012). Even within arguably 'pure' cases of cultural-familial ID, where intellectual impairment is mild and no additional symptoms or comorbid features are present, accumulating evidence indicates that intellectual functioning is in fact a polygenetic condition (Butcher et al., 2005; Kaufman, Ayub, & Vincent, 2010). To this end, the term nonspecific ID will hereafter replace the category of cultural-familial ID, and will refer to all cases of ID without diagnosed genetic or organic impairment.

Application of aetiological groupings in ESH research. Defining the underlying etiology of ID is relevant to this thesis insomuch that it informs our understanding of the developmental trajectories and associated behavioural phenotypes for various subgroups. Behavioural phenotypes describe characteristic cognitive and behavioural patterns associated with different types of disability. For example, the emotion specificity hypothesis is a statement about a behavioural phenotype (emotion recognition performance) of individuals with intellectual disability. The developmental path for individuals with a diagnosis of ID is not homogenous and therefore we cannot expect to extrapolate cognitive and behavioural outcomes from samples of mixed aetiology. This section makes a case for the continued inquiry regarding the ESH based on samples with distinct aetiological groupings, particularly nonspecific ID.

The developmental trajectory for individuals with nonspecific ID is best described with reference to the *similar sequence* and *similar structure* hypotheses. The two hypotheses were developed to explain the cognitive development of individuals with nonspecific ID, specifically the sequence and organisation of developmental constructs (Zigler, 1969). The similar sequence hypothesis builds on the basic assumption of developmental theory that the trajectory of cognitive functioning follows an invariant sequence, which by reason of its universality remains intact in those with ID, albeit somewhat slower with a lower eventual ceiling (Benson & Haith, 2009). Complementing the similar sequence hypothesis, the similar structure hypothesis is concerned with relationship across domains of functioning at a single point in time. The similar structure hypothesis proposes that individuals with ID demonstrate many of the same behaviours and underlying processes as neurotypical individuals who are at the same level of cognitive functioning. The similar structure hypothesis is aligned with Piaget's notion of an organised system, whereby developmental states are meaningfully related and unfold in a hierarchical manner (Burack, Russo, Gordon Green, Landry, & Iarocci, 2016).

The similar sequence hypothesis is largely supported across both organic and nonspecific ID populations (Weisz & Zigler, 1979), while studies on the similar structure hypotheses have produced more mixed results. The similar structure hypothesis is largely supported as it relates to those with nonspecific ID (Weisz & Yeates, 1981), the exception being some variation observed in respect to discrete domains of information processing (memory, learning set, and discrimination learning) (Weiss et al., 1986). However the implications of these findings are not definitive being that impaired performance was limited to task paradigms that that were long, repetitive, had limited ecological validity or were sensitive to the effects of motivational and personality factors (Burack, Russo, Gordon Green, Landry, & Iarocci, 2016). In contrast and as one might expect, research involving individuals with organic ID has yielded mixed results, dependent on the underlying pathology (Burack et al., 2016). These findings have led to a proliferation of syndrome-specific behavioural phenotype research, both in a general sense and emotion focussed investigations (e.g., Gagliardi et al., 2003; Whittington & Holland, 2011; Williams, Wishart, Pitcairn, & Willis, 2005). Syndrome-specific studies to date suggest diverse emotion recognition abilities in individuals with organic ID, whereby evidence in favour of the ESH holds for some populations (e.g., Down syndrome), and not for others (e.g., Fragile X) (Wishart, Cebula, Willis, & Pitcairn, 2007). Evidence for the ESH in nonspecific ID, reviewed in Chapter 3, remains less clear.

Aetiological parameters for the current thesis. In an already heavily stigmatised cohort, it is important to carefully examine additional hypotheses that may suggest further impairment, such as the emotion specific hypothesis. It must be appreciated that the ID diagnosis is an outcome state subject to a wide variety of causes which follow unique developmental trajectories. Accordingly, any attempts to combine aetiological groups may result in syndrome-specific patterns offsetting one another obscuring group specific strengths and weaknesses (Burack et al., 2016). Researchers should therefore remain wary of the blanket application of developmental hypotheses, such as the ESH.

Considering the wide support for the similar sequence similar structure hypotheses in nonspecific ID, it is particularly worthwhile to consider the legitimacy of the ESH in this group. If individuals with nonspecific ID supposedly follow a similar developmental trajectory and pattern to neurotypical individuals, then this raises the question - What is the underlying premise for the domain-specific emotion recognition deficit suggested by the ESH? Before the potential underlying mechanism leading to emotion recognition deficits can be delineated, further research is required to unequivocally demonstrate whether the ESH holds for nonspecific intellectual disability. While this group remains difficult to define, individuals in this group contribute to near half the population with ID (Maulik, Mascarenhas, Mathers, Dua, & Saxena, 2011). The scope of the current thesis is therefore limited to the investigation of nonspecific intellectual disability.

CHAPTER 2

CONCEPTUAL FRAMEWORK

Before one can measure emotion, one must define the phenomena. It is widely accepted that emotions may be characterised in terms of physiology (neural circuits), response systems, and feeling states that influence cognition and behaviour such as approach or avoidance. Emotions can also include antecedent cognitive appraisals (you have emotion about something), be social or relational in nature, include social-communicative signals such as facial expression, and may provide information regarding one's subjective feeling state (Izard, 2010). Which of these factors are necessary or sufficient in conceptualising the construct of emotion is subject to continued debate (Power & Dalgleish, 2016), and beyond the scope of this review. This chapter will narrow the field of enquiry to include two dominant emotion theories applied in experimental psychology, and how these inform the operationalisation of facial expression in emotion recognition paradigms and the current study.

What do facial expressions express?

Proper assessment of the ESH necessitates that facial expressions and emotional state be connected in a way that is both explicit and measurable. Several popular psychological theories exist to describe the theoretical basis of facial expressions, yet only two, discrete and componential appraisal theories, have offered testable predictions associating specific facial movements with different emotional states. Accordingly, only discrete and componential appraisal theories were considered in defining the theoretical framework for the current study.

Discrete emotion theories. Discrete emotion theories conceptualise emotions as clearly distinguishable categorical entities (Ekman & Cordaro, 2011). Discrete theories that remain under active development within quantitative research typically have their roots in Darwinism and are based on the idea that primary human emotions are innate and universal.

That is to say that they remain continuous across cultures, time and place (Darwin, 1998/1872). Consistent with evolutionary theory, facial displays of emotions are also predicted to serve a survival-critical function, either in regard to communicative purposes or physical functioning. For example, disgust causes the muscles above the lip to pull up and the nose to wrinkle, constricting the facial orifices away from potentially harmful or pathogenic objects (Darwin, 1998/1872).

Building on Darwin's notion of universal and innate emotions, Tomkins (1962) proposed that all humans experience a common set of *basic* phylogenetically evolved emotions that are expressed through specialised affect programs. The underlying assumption is that certain conditions, specifically different gradients of neural firing, automatically elicit a pattern of physiological responses which include muscular innervation (i.e., the basis of facial expression). This concept of *basic emotions*, also referred to as *first-order emotions* has been popularised by Ekman (Basic Emotion Theory) and Izard (Differential Emotion Theory), who together have refined Tomkins' (1962) original list of emotions to include five common emotions: happiness, sadness, fear, anger, and disgust (Ekman & Cordaro, 2011; Izard, 2011). In addition, Ekman also recognises surprise, and both researchers posit that contempt may be a basic emotion; however, Izard remains undecided citing a lack of clear-cut evidence (Izard, 2011). It is noteworthy, that Izard (2011) stresses that the incidence of first-order emotions decreases as normal development progresses, due to emotional expression becoming increasingly complex as feelings combine with higher-order cognitive processes.

Theorists in this tradition have tended not to elaborate on Tomkins' (1962) concept of specific neuromotor affect programmes, instead focusing on affect systems as 'open programmes' influenced by culture and learning. The described 'openness' is demonstrated in the individual and cultural variation (or culture specific 'display rules') seen in the initial regulatory patterns associated with different emotion categories. Such differences are attributed to environmental influence (Ekman & Cordaro, 2011). Put differently, the contributing circuits for 'basic emotions' remain prewired, albeit epigenetically moulded (Panksepp & Watt, 2011). This view accounts for both observed individual differences in facial behaviour and cases where prototypical components of a specific expression may not be expressed (Scherer & Ellgring, 2007).

Discrete emotion theories provide the dominant paradigm applied in quantitative research. Contributing hypotheses have received wide empirical validation, specifically, the genetic basis for discrete emotions (Kendler et al., 2008); consistent neural correlates connected with basic emotions (e.g., Hamann, 2012); discrete patterning of facial behaviour and bodily responses for 'basic' emotions (e.g., Nummenmaa, Glerean, Hari, & Hietanen, 2014; Young et al., 1997); independence of discrete categorisation of emotion from lexical categories (e.g., Sauter, LeGuen, & Haun, 2011); the early onset of discrete emotions in infants (e.g., Izard et al., 1995); and the universality of basic emotions based on genetic factors or evolved emotion-response systems rather than social learning, as demonstrated in studies with blind individuals (e.g., Hwang & Matsumoto, 2015; Matsumoto & Willingham, 2009).

Componential appraisal theories. Componential models of emotion typically adopt a functional approach in the Darwinian tradition, suggesting that emotional responses guide adaptive and flexible behavioural responses to important events (Grandjean & Scherer, 2008). Emotion is defined as the result of the appraisal or evaluation of antecedent events (see Moors, Ellsworth, Scherer, & Frijda [2013] for a brief overview and latest extensions to componential theories). The term appraisal is used to operationalise the significance of an event for an individual, and may occur at one or a combination of several levels of processing. For example, processing domains may include: basic stimulus characteristics such as novelty and pleasantness; motivational relevance as measured by needs, goals and values; secondary appraisals in the form of perceived ability to cope; or social dimensions characterised by identity, norms, values and justice (Ellsworth & Scherer, 2003).

In componential models elicitation of emotion is considered a dynamic process influenced by a variety of component processes, including cognition, action tendencies, physiological response, subjective feeling states, muscle innervation and subsequent motor expression. The patterning of the expression in each domain is the result of continuous and recursive evaluative process, whereby appraisal serves as both a causal influence and a component of emotion (Scherer & Ellgring, 2007). For example, a specific appraisal may lead to physiological changes and behavioural response which may then shape new appraisals.

Similar to discrete theories, componential theorists recognise that facial expressions represent different emotion states, though componential models typically endorse a larger number of highly differentiated emotion categories (Scherer & Ellgring, 2007). Nevertheless, level of emotion differentiation is variable, both across discrete and componential theories and within the componential camp. Lazurus (1991) offers a relatively conservative perspective, and proposes that the nature of emotion is subject to cognitive appraisal based on the importance of the event and an individual's ability to cope with it. Lazarus (1991) advocates a small number of fundamental appraisal groupings referred to as *relational themes* (e.g., 'loss of a cherished person or thing giving rise to sadness'), the consequence of which is the expression of a limited number of major emotions. While the focus here is on the specific elicitation of emotion, the notion re-joins several of the key assumptions of discrete emotion theories.

At the opposite end of the spectrum is Scherer's (1987, 1994) Component Process Model. The Component Process Model suggests that there are many highly differentiated emotional states, limited only by the available number of potential differential patterns of appraisal. In this paradigm, emotions are elicited from situations that have a direct bearing on an individual's goals, values and needs, and the subsequent subjective appraisal of their implication and consequences. Scherer (1987; 1994) also nods to a core set of emotions, 'modal' emotions, corresponding to commonly occurring appraisal patterns that result from universal organism-environment interactions (e.g., anger in reaction to blocked goals). Nevertheless, such labels capture only central tendencies of regularly recurring mental states and evaluation processes, and do not represent the 'readout' of motor programs as suggested by discrete theories. Remaining theorists in this tradition, typically adopt an intermediary view regarding the number of emotional states and their congruence with basic emotion theory (e.g., Roseman, Wiest, & Swartz, 1994; Smith & Scott, 1997). Evidence for componential appraisal theories is derived from research demonstrating systematic links between appraisals and emotion component responses, including but not limited to physiological response, (e.g., Gentsch, Grandjean, & Scherer, 2013; Grandjean & Scherer, 2008), some components of facial expression (e.g., Scherer & Ellgring, 2007), and action tendencies (e.g., Frijda, Kuipers, & ter Schure, 1989).

Implications for the measurement of facial expression

At first glance, claimed differences in emotion conceptualisations offered by discrete theories (e.g., basic emotions) and componential appraisal theories (e.g., relational themes) may seem an issue of semantics. However, these two theoretical perspectives have different implications regarding the measurement of emotion, specifically the precise facial muscle activity assigned to different emotional states and the temporal implications for the development facial expression stimuli.

The typical paradigm applied by discrete theorists to investigate emotion recognition, involves presenting participants with images or recordings of emotional expressions and asking them to describe the stimuli according to the basic emotion categories. Response options represent emotional states that in their simplest form are common to all people. To guide the accurate development of experimental stimuli, several facial coding systems have been developed. Examples include, A System for Identifying Affect Expressions by Holistic Judgment (AFFEX; Izard, Dougherty, & Hembree, 1983), the Maximally Discriminative Facial Movement Coding System (MAX; Izard, 1983), and the popular Facial Action Coding System (FACS; Ekman, Friesen, & Hager, 1978). The latter is a descriptive system designed to taxonomise facial behaviour on the basis of single facial muscles (action units). Action units may then be interpreted in the context of a secondary interpretive emotion coding system (e.g., Emotion FACS [EMFACS]; Friesen & Ekman, 1983).

Consistent with discrete theories, componential theorists apply categorical labels to describe the evaluative processes that result in various facial expressions. However, despite generally accepted themes (e.g., other blame leading to anger), appraisal has proven to be a partially subjective process and there is limited evidence suggesting a one-to-one to one relationship between appraisal content and specific facial movements (Nezlek, Vansteelandt, Van Mechelen, & Kuppens, 2008). Only a handful of facial action units have been neatly mapped onto specific appraisal themes (Mortillaro, Meuleman, & Scherer, 2012). Accordingly, specific facial movements associated with categorical emotion labels derived from this paradigm have observed considerable overlap (Scherer & Ellgring, 2007). Without well-validated predictions regarding the action units associated with different emotional states or appraisals, the replicable development of emotional stimuli for the purposes of emotion recognition research is not possible.

This issue is compounded by componential theorists' position on emotions as processes rather than states. Componential appraisal models have been developed to capture the richness of emotion differentiation which results from a process of constant change in all its component subsystems. Emotions are described as having an emergent quality, but it is the emergent nature of emotion that is difficult to operationalise in emotion recognition research. Scherer and Ellgring (2007) suggested that the dynamic coding of facial behaviour is required to empirically test the sequential process inherent in the Component Process Model. Such research is a prerequisite for the valid production of emotion stimuli for the purposes of recognition research. In contrast to the componential theories, discrete theorists predict that emotions are triggered as a package, where prototypical affect programmes are initiated with a stimulus and increase in intensity before returning to baseline. The notion of a prewritten 'programme' implies that the production of basic emotion stimuli for experimental purposes should be produced in a smooth uninterrupted flow (Scherer & Ellgring, 2007).

As is often the case in psychological research, both classes of theory hold merit and demonstrate theoretical and empirical support. Discrete theories of emotion have been selected to inform the current thesis based on consideration of the different structures of emotion offered by the two theories, and the implications for developing a replicable emotion recognition task. Compatibility with previous research was also considered. Discrete theories serve as the dominant paradigm in quantitative psychological research of emotion. In a recent study including 248 established emotion scientists, when asked to identify their theoretical orientation 49% identified discrete emotions and only 11% identified "emotions as constructed, either socially or psychologically to fit current conditions" (Ekman, 2016, p. 32). The latter description consistent with the basic premise of componential theories. The strong favour of discrete models is also apparent in the emotion specificity literature (see Scotland, Cossar, & McKenzie, 2015 for a review) and forms the typical basis for experimental research into facial emotion processing as it is described in the following chapter. In the interests of interpreting findings from the current study in the context of previous research, discrete theories offer a natural starting point and provide the framework for this thesis.

CHAPTER 3

THE EMPIRICAL BASIS OF THE EMOTION SPECIFICITY HYPOTHESIS

Facial emotion recognition ability has been studied across a wide range of neurodevelopmental disorders (e.g., attention-deficit/hyperactivity disorder [Bisch et al., 2016]; Down syndrome [Pochon & Declercq, 2013]; Williams syndrome [Gagliardi et al., 2003]; Fragile-X syndrome [Turk, Cornish, & Cornish, 1998]), with close attention being paid to autism spectrum disorder (Uljarevic & Hamilton, 2013). Individuals with nonspecific intellectual disability have received less attention in this regard, despite being the largest single group within the ID population (Wishart, Cebula, Willis, & Pitcairn, 2007). While there is some evidence that this group may also struggle with decoding or interpreting emotional facial cues (Scotland et al., 2015; Zaja & Rojahn, 2008), the mechanisms behind observed performance deficits in emotion-based laboratory tasks have yet to be delineated. The fundamental question remains: are observed difficulties secondary to the cognitive limitations inherent in intellectual disability, or better explained by the emotion specificity hypothesis (ESH; Rojahn, Rabold, & Schneider, 1995) as domain-specific facial emotion processing deficits?

In this chapter, a case is made for continued enquiry into the ESH in response to the lack of clear-cut evidence. Minimal criteria for testing the ESH are established and provide context for a review of the literature. An alternative explanation to the ESH is also presented, emphasising limitations in existing facial recognition task paradigms. Finally, available emotion recognition measures are reviewed highlighting the need for a new measure in the interests of enhancing ESH research.

Empirical basis of the emotion-specificity hypothesis in intellectual disability

Emotion recognition research can be organised according to three levels: biology, cognition and behaviour. Biological approaches typically draw on brain activation or lesion studies to localise emotion related neural structures, while cognitive approaches consider how emotion stimuli are processed (e.g., featural versus configural processing of facial cues). The behavioural approach focuses exclusively on task performance and is most recognisably aligned with real world implications and clinical practice. This review focuses on behavioural research, with brief context provided in terms of the biological and cognitive underpinnings of emotion recognition.

Recognising emotion from facial expressions involves a two stage processes subserved by a wide range of neural structures. The first involves the *perception* of changes in facial behaviour, and the second, the assignment of emotional significance and subsequent emotion *recognition* (Adolphs, 2002). The early perceptual processing of faces draws on many of the same brain structures involved in general visual processing, in addition to the sensory cortices of the occipital and temporal lobes responsible for more fine-grained representations of facial features and their configuration (Adolphs, 2002). There is also some evidence that facial expressions are processed separately from facial identity, based on a key double dissociation observed in disorders with established socio-emotional processing difficulties (Bate & Bennetts, 2015). Perceptual processing occurs independent of higher level cognition and depends solely on the geometric properties provided by facial stimuli. In contrast, recognition is dependent on knowledge of the emotion signalled and the ability to initiate this knowledge base in response to perceptual cues. Recognition of facial emotion draws on regions of the amygdala and orbitofrontal cortex (Adolphs, 2002).

Informed by the processes underlying emotion recognition, Rojahn and Zaja (2007) proposed the following minimal criteria to test the ESH: two participant groups, an experimental group with ID and a typically developing mental age matched control group; combined with an experimental task of emotion processing and a control task of comparable abstraction and complexity. Inclusion of a control task devoid of emotional content, provides insight into the specificity of observed deficits, while mental age matching controls for IQ related group differences tapped by the task on which participants are matched (Moore, 2001). Based on these criteria, full support for the ESH requires that individuals with ID demonstrate an exclusive deficit on tasks tapping emotion recognition, and comparable performance to control subjects on control measures. A review of the literature identified only six studies (Hobson, Ouston, & Lee, 1989a; Hobson, Ouston, & Lee, 1989b; Rojahn, Rabold, et al., 1995; Scotland, McKenzie, Cossar, Murray, & Michie, 2016; Williams et al., 2005; Wishart et al., 2007) that met both the minimum criteria set forth by Rojahn and Zaja (2007) and included an experimental group with nonspecific ID. Studies that did not explicitly state an aetiological basis for diagnosis were included if they did not intentionally combine individuals with ID related syndromes or known organic causes. Table 3.1 presents a summary of the six studies, including participant characteristics, emotion stimuli type, emotions examined, control tasks and findings.

The earliest study, completed by Hobson and colleagues (1989a), included adolescents and young adults with mild intellectual disability and a verbally mental age matched control group. The experimental task involved matching emotionally expressive voices to a corresponding black and white photograph of an emotionally expressive face portraying one of six emotions (happiness, unhappiness, anger, fear, surprise and disgust). For the control task, participants were required to select a picture of an object to accompany a recorded sound (e.g., images and audio recordings representing different types of motor vehicles). For both tasks, respondents selected their answer from a series of six distractor stimuli. Consistent with the ESH, the two groups performed comparably on the control task while control subjects outperformed individuals with ID on the emotion matching task. Regrettably, the control task in this study did not include facial stimuli and therefore the results do not differentiate between a general deficit in the processing of faces and an emotion-specific processing deficit.

In a second study including the same participants and a subset of the original stimuli, the relative performance deficit for facial versus non-facial stimuli was no longer observed when participants were allowed to freely label stimuli rather than using cross-modal matching (Hobson et al., 1989b). Instead control participants were found to outperform individuals with ID on both tasks, indicating that results of the previous study may have been an artefact of an overly complex task paradigm rather than evidence of a facial emotion processing deficit. Cross-modal matching requires the respondent to simultaneously attend to and remember phonological information, access meaning across modalities, and select a response amongst several distractors. One explanation is that these processes, specifically the differentiation of distractor stimuli, would be more difficult for facial emotion than for non-emotional objects, given that visual differences across emotion-stimuli are arguably more subtle than those observed within object categories (e.g., vehicles).

Rojahn, Rabold, et al. (1995) improved on issues of task complexity and specificity. Participants in the experimental group included 16 adults with mild to moderate intellectual disability. Again, the control group was matched on verbal ability and experimental stimuli included black and white photographs. Participants were required to categorise facial stimuli as 'happy', 'sad' or 'happy nor sad', and 'young', 'old', or 'young nor old', for the index and control tasks respectively. Stimuli were then further differentiated on a five-point scale along the dimensions of 'a little' or 'a lot'. Consistent with the ESH, based on task total scores individuals with ID performed comparably with mental age matched controls on the age discrimination task and were significantly less accurate on the emotion-based task.

Findings from the Rojahn, Rabold, et al. (1995) study have served as a cornerstone in the ESH debate. However, further exploration of the data revealed alternative explanations that have yet to be explored in subsequent research: specifically, the source of performance differences as they relate to emotion category. No statistically significant differences were observed between individuals with ID and mental age matched controls on items related to happiness. Further, differences between the two groups on sadness related items were arguably minor. While a statistical difference in favour of mental age matched controls was reported for sadness items, curiously, mean data presented pictorially indicated that the ID group performed marginally better. Reported discrepancies in the direction of descriptive data and statistically tested group differences are likely a result of the within matched pair variance controlled for in the inferential analysis. Nevertheless, what is noteworthy here is the small magnitude of the observed group difference, particularly given the context of the modest sample size. What then is the main cause for the observed deficits for participants with ID? The largest group differences were observed for emotionally neutral items represented by the 'happy nor sad' category. Until it is determined why rating neutral expressions is particularly challenging for people with ID, the contribution of this study to the ESH debate remains limited, in that the supporting evidence largely rests on emotion recognition performance for faces with no emotional content.

Two additional syndrome-specific studies have also inadvertently offered insight regarding the ESH, by way of including individuals with nonspecific ID as control participants. Williams, Wishart, Pitcairn, and Willis (2005) and Wishart et al. (2007) explored facial emotion recognition skills in children and adolescents with nonspecific ID and typically developing mental age matched children. Performance measures included an emotion-matching (happiness, sadness, anger, disgust, surprise and fear) to sample task and an identity-matching task to control for basic face processing. Participants were required to select responses from three black and white photographs. Neither study reported results in favour of the ESH. Williams et al. (2005) found that individuals with ID did not perform as well as mental age matched control subjects on the emotion-matching task, yet outperformed control subjects on the identity-matching task. Further both groups observed superior performance on the emotion-matching task when compared with the identity-matching task, with the control group demonstrating a larger total score discrepancy across the two. At the time of publishing, the observed differences were not substantiated by statistical tests. Fortunately, in a subsequent syndrome-specific study using partial data from Williams et al. (2005), Wishart and colleagues (2007) applied inferential analysis to reveal no statistically significant group performance differences across both the index and control tasks.

In contrast, a recent study by Scotland et al., (2016) has offered comparatively robust evidence in favour of the ESH. Participants included adults with intellectual disability recruited via community and forensic services, and a control group matched on estimated cognitive ability. The experimental task involved three conditions; emotion naming, and emotion recognition from a choice of nine or two emotion labels. Stimuli included nine emotions (happy, sad, afraid, angry, bored, worried, surprised, disgusted and neutral) in the form of line drawings, and colour photographs of the face in isolation and with context. Control tasks provided equivalent forms of the index tasks but without emotional content, and participants were required to identify visual features such as hair or eye colour. Consistent with the ESH, control participants outperformed participants with ID across all emotion recognition tasks and stimuli types.

The specificity of observed deficits in the Scotland et al., (2016) study remain less clear, due to control participants outperforming individuals with ID on the control task. Both experimental groups performed better on the control tasks than the emotion tasks, with participants with ID observing a larger performance discrepancy. Regrettably, within group pairwise comparisons were not reported due to ceiling effects on the control tasks by the control group, and it remains unknown whether the observed differences were statistically significant for one or both experimental groups. Should differences have been significant for both groups, rather than offering supporting evidence for the ESH this would suggest that the experimental tasks were not an appropriate match in terms of cognitive demands. Arguably a very real possibility, given that the identification of emotion (which includes a combination of subtle facial cues) is not comparable in level of abstraction to the identification of basic visual features (such as hair colour). Further, replication of this study is particularly important due to a large proportion of the participants having been recruited from forensic services. Significant impairment in the ability to identify facial expressions has been documented in the general offending population (e.g., Hoaken, Allaby, & Earle, 2007; Robinson et al., 2012). Therefore, it may be that the difficulties observed by individuals with ID are independent of their disability, or at least in part, better explained by the mechanisms linking offending and emotion recognition deficits in the general population.

Table 3.1

Summary of Studies Meeting Rojan and Zaja's (2007) Minimum Criteria

	Experimental group					Task performance		
Authors	Intellectual Disability (ID)	Typically developing control group (TD)	Emotions assessed	Stimuli	Experimental tasks	Emotion task	Control task	Evidence of specificity (group x task interaction)
Hobson, Ouston, & Lee (1989a)	Adults and adolescents 16 male, 5 female n=21 CA: 12.5-25.8 (*M*=18.4) MA: 4.1 to 11.0 years (*M*=7.01)	Children 13 male, 8 female n=21 CA: 4.8-11.6 (*M*=7.2) MA: 4.1-10.1 (*M*=7.0)	Happy Unhappy Anger Disgust Fear Surprise	Black and white photographs (emotionally expressive faces and objects) Audio recordings (emotionally expressive voices and non-emotional sounds)	Picture-sound matching	ID<TD	ID=TD	Yes
Hobson, Ouston, & Lee (1989b)	Same as above	Same as above	Same as above	Same as above	Picture labelling	ID<TD	ID<TD	No
Rojahn, Rabold, & Schneider (1995)	Adults n=16 CA: not reported MA: IQ 40-70	Children N=16 CA: 6.5-12 years (*M*=not reported) MA: not reported Adults n=16 CA: not reported (older than index group)	Happy Sad Neutral	Black and white photographs (emotionally expressive faces and people of different ages)	Category recognition (emotion and age) Rating (emotion intensity/extent of age group membership)	ID<TD ID<TD	ID=TD ID<TD	Yes

Table 3.1 Continued

Authors	Experimental group		Emotions assessed	Stimuli	Experimental tasks	Task performance		
	Intellectual Disability (ID)	Typically developing control group (TD)				Emotion task	Control task	Evidence of specificity (group x task interaction)
Williams, Wishart, Pitcairn, & Willis (2005)[a]	Children and adolescents n=53 29 Male, 24 female CA: 6-17.4 (*M*=11.83 (*SD*=38.1) MA: 3.1-6.0 (*M*=4.67, *SD*=9.8)	Children n=39 22 male, 17 female CA: 2.8-5.6 (*M*=4.0, *SD*=9.2) MA: 3.1-6.0 (*M* = 4.4, *SD*=10.3)	Happy Sad Angry Disgust Surprise Fear	Black and white photographs (emotionally expressive faces)	Matching to sample (emotion and identity) from a choice of three	ID<TD†	ID>TD†	Not tested
Wishart, Cebula, Willis, & Pitcairn (2007) [a]	Children and adolescents n=15 CA: 6.0-18.1 (*M*=11.9, *SD*=3.4) MA: range not reported (*M*=4.2, *SD*=1.1)	Children n=15 CA: 2.8-7.6 (*M*=4.0, *SD*=1.17) MA: range not reported (*M*=4.1, *SD*=1.1)	Happy Sad Angry Disgust Surprise Fear	Black and white photographs (emotionally expressive faces)	Matching to sample (emotion and identity) from a choice of three	ID=TD	ID=TD	No
Scotland et al. (2016)	Adults n=23 18 male, 5 female CA: 25-61 (*M* = 45.7, *SD*=11.3) MA: estimated IQ *M*=68.3 (*SD*=7.7)	Children n=23 11 male, 12 female CA: 7-13 (*M*=10.4, *SD*=1.7) MA: estimated adult IQ equivalent *M*=74.2 (*SD*= 9.4)	Happy Sad Afraid Angry Bored Worried Surprised	Line drawings and colour photographs with and without context (emotionally expressive faces)	Emotion/feature labelling Emotion/feature recognition from a choice of nine and two	ID<TD	ID<TD	Not tested

Note. CA = Chronological age, MA = Mental age, *M* = Mean, *SD* = Standard error of the mean.
† Reported differences not tested for statistical significance in original article.
[a]. Experimental groups not pertinent to the testing of the ESH have not been included.

In summary, when we consider the current evidence base against Rojahn and Zaja's (2007) minimal requirements for testing the ESH, we are left with mixed results. Four of the six reviewed studies provided evidence to support the hypothesis in a roundabout way (Hobson et al., 1989a; Rojahn, Rabold, et al., 1995; Scotland et al., 2016; Williams et al., 2005); however, the validity of these findings remains questionable. In all but one instance where group differences were not verified by statistical tests (Williams et al., 2005), the validity of the findings are complicated by a combination of highly complex task demands, absence of within group statistical comparisons to determine the specificity of observed deficits, discrepant task related demands including level of abstraction across index and control tasks, and participant related confounds. Further, Rojahn, Rabold, and Schneider's (1995) landmark study, which is largely free of the listed methodological limitations, found that emotion recognition difficulties experienced by individuals with ID were mostly limited to neutral expressions. While it would be premature to discount the ESH at this stage, research supported by inferential statistics with the proper task and participant controls is required if we are to determine its true relevance in intellectual disability.

An alternative explanation to the emotion specificity hypothesis

The basic premise behind the ESH has evolved from top-down reasoning, the rationale based on findings in the general population linking emotion recognition ability to successful social integration and any type of learning based on social consequences (Rojahn, Rabold, et al., 1995). The assumption is that in the case of ID, where social adaptive difficulties are an integral part of the diagnosis, individuals must manifest domain specific deficits in recognising facial emotions. While this explanation is plausible, there exist a number of alternative ID related factors that negatively impact social functioning independent of emotion recognition capacities (e.g., institution versus community dwelling [Hetzroni & Oren, 2002]).

An alternative hypothesis has been put forward by Moore (2000), who proposed that facial emotion perception capabilities are intact in people with ID and that poor performance on emotion recognition tasks may be accounted for by reduced IQ-related information-processing abilities. Moore's (2000) idea is based on reports from ecological psychologists who suggest that as a function of natural selection, humans have developed a species-level preparedness to encode stimuli relevant to social interaction. This includes the development of separate neural structures dedicated to the encoding of facial stimuli and objects (Kennerknecht, Kischka, Stemper, Elze, & Stollhoff, 2011). Consequently, facial processing capabilities need not be constrained by general cognitive abilities (Moore, 2001). Consistent with this rationale, a study by Dobson and Rust (1994) showed that relative to typically developing mental age matched controls, adolescents with ID who experience deficits encoding and remembering objects had spared capacities for remembering and encoding faces.

Why then are typically developing controls consistently found to outperform individuals with ID on emotion recognition tasks? Moore (2000) suggests that domain specific emotion recognition difficulties may be overstated in the literature, due to existing emotion recognition paradigms making extensive cognitive demands that disadvantage people with ID. There appears to be some evidence of a positive relationship between the information-processing demands of emotion recognition tasks (e.g., memory, attention and abstraction) and performance discrepancies observed between individuals with ID and mental age matched controls. Even basic experimental tasks require participants not only to attend to stimuli, but also encode and discriminate between stimuli before offering a response (Moore, 2000). Therefore, a meaningful interpretation of participant performance must differentiate between participant capacities for emotion recognition and their ability to cope with task specific demands beyond those required to perceive emotional facial cues. As mentioned earlier, Rojahn and Zaja (2007) suggest that this may be achieved by including a mental age matched control group and at least one carefully designed control task of comparable complexity and abstraction. While more than twenty studies claim to offer relevant information in favour of the ESH (for review see Moore, 2001; Rojahn, Lederer, & Tassé, 1995; Rojahn & Zaja, 2007; Scotland et al., 2015) only the six previously reviewed studies met the minimal criteria, and even across these few cases findings were mixed.

Whether the remaining studies offer a meaningful contribution to the ESH remains debatable. Accordingly, one must remain cautious in concluding that it is the emotional content of stimuli that is responsible for reported deficits observed by individuals with ID.

In addition to the small number of available studies controlling for task related demands, the issue is further confused by limitations in the ecological validity of commonly used facial emotion stimuli. Significant limitations in ecological validity, at a minimum compromise the legitimacy of existing findings and in more severe cases potentially bias research in favour of the ESH. If stimuli used to tap emotion recognition abilities are not natural representations of human emotion, then performance deficits observed in individuals with ID may be due to any number of neural systems involved in processing the selected stimuli. Performance on these tasks may be independent of emotion recognition capacities and may rely on more general aspects of intelligence (Moore, 2001).

These criticisms are most relevant to studies which employ emotion stimuli in the form of line or schematic drawings (e.g., Brosgole, Gioia, & Zingmond, 1986; McKenzie, Matheson, McKaskie, Hamilton, & Murray, 2001; Scotland et al., 2016). Such stimuli do not provide a close approximation of real-world emotions, and therefore it cannot be assumed that performance deficits represent emotion perception or recognition abilities. Rather participant performance may be more representative of cognitive inferential and classification abilities, or a learned culturally influenced "emotional shorthand" (Moore, 2001). Further, neurophysiological studies using functional magnetic resonance imaging have found that cells in the fusiform face area (cells preferentially tuned to facial identity) are more sensitive to photographic stimuli relative to schematic representations of faces (Tong, Nakayama, Moscovitch, Weinrib, & Kanwisher, 2000). This suggests that the brain processes real facial stimuli in a quantitatively different manner to schematic representations, and therefore the real-world implications of findings derived from schematic stimuli may be limited.

Even the use of photographic stimuli, which on the surface appear to offer a close approximation to real-world emotions, have considerable limitations. The most significant criticism is that photographs neglect the subtle temporal aspects of emotion. For example, expressions of sadness typically emerge slowly, while happiness is associated with rapid facial movements (Kamachi et al., 2001). A recent review of behavioural, brain lesion, facial electromyography and neuroimaging studies concluded that dynamic cues, such as those captured by video, promote more accurate emotion recognition when compared with static images (Alves, 2013). While the specific mechanisms remain unknown, Moore (2001) speculates that a static image fails to capture the complex moving configuration of facial features, and that the dynamic change of the relations among the features over time may contribute to the accurate recognition of emotion. Consistent with this notion are reports that the extra information supplied using multiple static images (or 'morphing' techniques) does not lead to increased accuracy, suggesting that motion enhances emotion recognition by showing *the way* expression has changed (e.g., Ambadar, Schooler, & Cohn, 2005; Bould & Morris, 2008). Further, neuroimaging and brain lesion studies suggest a dissociation between the neural substrates responsible for processing dynamic and static expressions (e.g., Kessler et al., 2011; Kilts, Egan, Gideon, Ely, & Hoffman, 2003). Finally, electromyography studies show that dynamic expressions are linked to greater changes in physiological responses associated with emotion recognition processes, specifically higher levels of facial mimicry displayed by the observer (e.g., Rymarczyk, Biele, Grabowska, & Majczynski, 2011; Sato & Yoshikawa, 2007). These findings describe key differences in how the brain processes static and dynamic images, highlighting that static images do not offer an ecologically valid approximation for real-world emotion.

Despite wide criticism directed to the use of static stimuli in general emotion research, the importance of facial dynamics is largely neglected by researchers in their assessment of the ESH. Intellectual disability is often accompanied by difficulties in a wide range of cognitive domains (e.g., working memory and processing speed; Koriakin et al., 2013) that could potentially affect how dynamic stimuli are processed, yet not one of the six previously reviewed studies included dynamic stimuli. The most common stimuli set applied in existing research remains the popular black and white photographs included in the Pictures of Facial Affect (Ekman & Friesen, 1976). It remains to be seen whether the effects of dynamic cues on emotion recognition performance are mediated by ID. To the author's knowledge, there exists only one study that explores this issue. Harwood, Hall and Shinkfield (1999) investigated the role of movement using 10-second video clips of the six basic emotions, and corresponding static images taken at the apex of the video clips. It was found that dynamic cues enhanced overall performance for both groups. However, group level effect sizes were not reported for the static and dynamic conditions, and it is unclear whether the advantageous effects of dynamic cues benefitted individuals with and without ID equally. Further, the power of the study was limited due to only 12 participants being included in each group. Replication is required to draw firm conclusions regarding the impact of dynamic cues, or lack thereof, in the existing ESH literature.

A final factor not considered in Moore's (2001) review, and largely neglected in the ID literature, is emotion intensity. There is some evidence that emotion recognition accuracy is contingent on the strength or intensity of the emotion, with higher intensity emotions being recognised more easily (e.g., Hess, Blairy, & Kleck, 1997; Hoffmann, Kessler, Eppel, Rukavina, & Traue, 2010; Montagne, Kessels, De Haan, & Perrett, 2007). These findings have not been corroborated with groups with ID, with some studies suggesting that the effects of intensity may differ from the general population. Specifically, Gray, Fraser, and Leudar, (1983) reported that individuals with mild ID do not benefit from increasing intensity, while individuals with severe ID may become confused with high intensity emotion.

The intensity of expression has also been observed to moderate the emotion recognition benefits of dynamic displays in the general population, where the benefit of motion cues is reduced with higher intensity expressions (Bould & Morris, 2008). These are particularly significant findings when we consider that the effects of dynamic cues remain largely untested for individuals with ID, in addition to the lack of verification of emotion intensity in existing studies. Further, amongst others (e.g., Wingenbach, Ashwin, & Brosnan, 2016), it is the author's subjective opinion that many of the stimuli applied in the previously reviewed ESH studies include exclusively high intensity emotions. If Gray, Fraser, and Leudar's (1983) findings reflect the true abilities of individuals with ID, it may be that reliance of existing ESH studies on high intensity expression presents an inflated view of impairment for this group.

If we consider the alternative explanation offered by Moore (2000) in the context of the methodological limitations outlined, it is the author's opinion that Rojahn and Zaja's (2007) minimal criteria for the assessment of the ESH is necessary but not sufficient to constitute a fair test. There is ample evidence to suggest that the common methodological practices that may be taken for granted in typically developing populations may disadvantage individuals with ID creating a bias in favour of the ESH. If we are to truly test the relevance of the ESH to individuals with ID, Rojahn and Zaja's (2007) minimal requirements should be met with additional recommendations for dynamic naturalistic emotion displays of varying intensities, and the careful control of task related demands. Regrettably, there has been a lack of interest in the development of an ecologically sound construct validated measure capable of differentiating individual emotion recognition capacities in people with ID.

Available facial emotion databases and measures

To the author's knowledge, there exist only four tests or expression databases that include facial representations of the universal emotions that are also presented; in colour, with dynamic cues (without additional bodily or auditory cues), and include a validated range of emotion intensities. Note that the additional criteria of colour stimuli has been included in this overview of quality measures and stimuli sets. Emotional states change the perfusion and oxygenation of blood under the skin, which cause subtle changes in skin colour to become visible (Kret, 2015). Presenting emotions in colour rather than grayscale, has been shown to improve emotion recognition (Zak, Laeng, & Simon-Liedtke, 2015) and effect the perceived intensity of the emotion (Barr & Kleck, 1995), which in turn may influence the specific emotion that is recognised (Hess et al., 1997). While there is no obvious reason why colour cues would impact the outcome of paradigms designed to test the ESH, they have been included as a criterion in the interest of ensuring ID specific research maintains the same quality standards afforded the general population.

The four tests or stimuli sets to be reviewed include the Emotion Recognition Task (ERT; Montagne, Kessels, De Haan, & Perrett, 2007), the Max-Planck Institute Facial Expression Database (MPI; Kaulard, Cunningham, Bü Lthoff, Wallraven, & Ernst, 2012), the Database of Facial Expressions (DaFEx; Battocchi, Pianesi, & Goren-Bar, 2005) and the Amsterdam Dynamic Facial Expression Set – Bath Intensity Variations (ADFES-BIV; Wingenbach, Ashwin, & Brosnan, 2016). The ERT is based on video sequences derived from morphed images, and the remaining three databases, true recordings of human faces.

The ERT includes colour images of six of the universal emotions (anger, sadness, disgust, happiness, surprise, and fear) and offers a wide range of emotional intensities (20%-100%). The limiting factor for the ERT however, is its use of morphed images. Morphed stimuli are developed by transforming one static image into another, usually a neutral expression and an emotional expression, by inserting a series of computer generated images along predefined linear increments (Montagne, Kessels, De Haan, & Perrett, 2007). Morphed stimuli allow for a high level of precision and standardization in respect to exposure times, but are limiting in experimental research in that forced stimuli changes undermine the naturalness of facial expressions - specifically, the activation of specific facial action units, and the timeframe in which they reach apex. Alteration of the natural temporal characteristics of emotion has been shown to reduce the perceived naturalness of displays and diminish recognition accuracy (Bould, Morris, & Wink, 2008; Kamachi et al., 2001).

In contrast, true video recordings, as demonstrated in the MPI, DaFEx and ADFES-BIV, preserve emotion specific variation in the onset and speed of facial action units. The MPI includes 55 facial expressions, including five basic emotions at two intensity levels. Unfortunately, published validation data is limited to stimuli of only five of nineteen models and includes only high intensity expressions. The specific intensity levels of emotional displays included in the MPI have also yet to be validated. Similarly, the three intensity levels (high, medium and low) offered by the DaFEx have yet to be corroborated by research; however inter-rater reliability data supports the individual emotion categories included in the measure (happiness, sadness, surprise, disgust, anger, fear and neutral).

The most significant limitation of the measures and stimuli sets described thus far is the absence of any replicable and objective measure to evaluate the legitimacy of emotion displays. All three measures relied on consensus scoring by lay persons. Based on the underlying assumption of discrete emotion theories concerning the evolutionary and social foundations of emotional expression (Ekman & Cordaro, 2011), consensus scoring is an appropriate tool to assess the validity of emotion displays. However, in isolation, the utility of consensus scoring in assessing items of lower intensity is limited. This is for the simple reason that it is not possible to differentiate between low scores caused by inaccurate emotional displays, and high-quality low-intensity items, which are more difficult to recognise due to the subtle nature of facial changes.

The creators of the ADFES-BIV remedied this issue by taking already FACS (Ekman, Friesen, & Hager, 1978) verified footage and then validating intensity ratings based on the response accuracy and latency of non-expert participants. It was assumed that response latency would be linked to the ease at which the emotion was recognised, with higher intensity emotions being easier to identify. The ADFES-BIV was published during the data collection phase of the current study, and to the author's knowledge is the only database available that includes objectively validated emotional displays of varying intensity, with colour stimuli and dynamic cues. The ADFES-BIV has not yet been applied to individuals with ID and further research is required to determine if items are geared at an appropriate level to discriminate between the emotion recognition capacities of individuals in this group.

All four of the described stimuli sets and tests offer an improvement on the typical static images that dominate ESH research. Nevertheless, these measures were not developed specifically in the interest of the nonspecific ID population and the described limitations make a case for the development of a new assessment measure. The benefits of a validated measure when compared with large stimulus sets cannot be overstated. The application of stimulus sets limit researchers to selecting a small pool of items from a larger set, which makes comparisons across studies difficult. The alternative option is then to apply all items in a set. The application of a large number of items does not lend itself to the study of individuals with limitations in basic cognitive abilities, as is the case in ID. Further in terms of clinical utility, the development of a measure validated specifically for use with those with ID will ensure stimuli difficulty is set at an appropriate level. This in turn would lead to improved observation of within group differences, and in time, the potential for accurate assessment of change following intervention and the development of group norms.

CHAPTER 4

EMOTION PERCEPTION AND SOCIAL ADAPTIVE FUNCTIONING

Determining the integrity of emotion recognition in nonspecific ID is important to our theoretical understanding of social adaptive functioning in this group. Deficits in social adaptive functioning are a defining criterion of intellectual disability. Component skills include social judgement, the ability to empathise and to initiate and maintain friendships, and an awareness of others' feelings, thoughts and experience (American Psychiatric Association, 2013). The proper development of these skills allow people to avoid interpersonal conflict (Matson & Swiezy, 1994) and benefit from healthy relationships (Bielecki & Swender, 2004). Regrettably, intellectual disability is associated with varying degrees of deficit in social adaptive functioning, which serves to negatively impact quality of life through social isolation, increased incidence of mental health difficulties and stigmatization (Matson & Hammer, 1996). Deficits in facial emotion recognition are often cited as a key component skill responsible for the social adaptive difficulties experienced by people with ID (Rojahn, Rabold, et al., 1995), yet few studies have attempted to demonstrate this relationship. This chapter provides a brief theoretical account for the emergence of social adaptive behaviour. The ESH is also considered in the wider context of social adaptive functioning and evidence connecting emotion recognition deficits and social behaviour reviewed.

Maladaptive or challenging social behaviour falls outside the scope of this review. While commonly perceived as extremes of the same continuum, social adaptive functioning and maladaptive social behaviour are distinct constructs and unlike social functioning, maladaptive behaviours are not included in the diagnostic criteria for ID (Schalock et al., 2012). Preliminary research suggests that maladaptive and adaptive social behaviours are only weakly related in individuals with ID, with moderate to strong relationships limited to individuals with comorbid conditions such as autism spectrum disorder (Tassé, 2009).

Social adaptive functioning defined

The field of social and neuropsychology is replete with terms describing social functioning and its component skills, many of them used interchangeably. For current purposes *social adaptive functioning* represents an overarching competency, informed by an individual's *social skills* and *social cognition*. In this paradigm, social cognition refers to the mental operations required to perceive and process interpersonal social cues and plan suitable responses (Yager & Ehmann, 2006). Social cognition is considered specialised or domain specific, meaning that while this capacity is dependent of requisite cognitive abilities such as simple attention, social cognition likely accounts for additional independent variance in social functioning (Jones & Day, 1997; Marlowe, 1986). The construct of social skills encompasses social cognition and includes the additional capacity to execute appropriate goal-directed social behaviour (Bedell & Lennox, 1997).

Numerous theoretical models have been proposed to account for the development of social adaptive functioning. Among the most influential, the reformulated Social Information Processing model (SIP; Crick & Dodge, 1994) focuses on the mechanism by which discrete elements of social cognition generate competent social behaviour. The SIP proposes that when faced with a social situation, individuals engage in a series of steps that are informed at each stage by the individual's existing social knowledge, schemas, and memories of previous social encounters. Processing steps include the encoding (i.e. sensation, perception, attention and focus) of external and internal cues, interpretation of these cues, and the generation, selection and enactment of goal related behaviour. The model posits that due to the complexity of social situations individuals are likely to engage in multiple processing steps in parallel. Notably however, the progression from a single stimulus (such as a disapproving look) to a behavioural response (such as withdrawal) follows the described steps in chronological sequence (Crick & Dodge, 1994).

Social adaptive functioning and the emotion specificity hypothesis

The SIP model proposes that adaptive social behaviour is dependent on the efficient execution of each of the described stages of processing. From a clinical and research perspective, when social difficulties are observed such as those associated with ID, the chronological sequence described by the SIP model allows the source of social difficulties to be isolated, providing opportunity for increasingly targeted intervention (Beauchamp & Anderson, 2010). The ESH is relevant here, in that it offers a compatible explanation for the cognitive processes responsible for the social difficulties observed by those with ID; specifically, a domain specific impairment in the encoding and interpretation of facially expressed emotion. It is sensible to focus research on these more rudimentary stages of social information processing, before investigating progressively more complex skills dependent on the additive effects of processing at the earlier stages.

Facial expressions offer insight into internal emotional states that can predict the intentions and reactions of the observed individual (Elfenbein, Foo, White, Tan, & Aik, 2007). Based on the SIP model, it may be anticipated that specific deficits in such a fundamental skill as emotion recognition ability, would likely contribute to broad reaching difficulty navigating one's social environment. Consequently, emotion recognition has been hypothesised as a requisite skill for normal social-adaptive development (Rojahn, Rabold, et al., 1995). Accordingly, several authors have recommended that social intervention programmes should focus on enhancing emotion recognition skills (Mcalpine, Singh, Ellis, Kendall, & Hampton, 1992; Owen, Browning, & Jones, 2001). Further, there exists accumulating evidence that facial emotion recognition is indeed a trainable skills in adults with ID (for a review see Wood & Kroese, 2007). However, such research may be premature, as studies demonstrating the link between facial emotion recognition ability and social adaptive functioning are limited. To the author's knowledge only six studies have directly addressed this issue (García-Villamisar, Rojahn, Zaja, & Jodra, 2010; Rojahn, Esbensen, & Hoch, 2006; Rojahn & Warren, 1997; Simon, Rosen, Grossman, & Pratowski, 1995; Williams et al., 2005; Wishart et al., 2007).

The most recent and psychometrically robust study, produced by García-Villamisar et al., (2010), assessed emotion recognition according to performance on two tasks: an emotion picture to picture matching task and a free picture labelling task. Social domains assessed relevant to this study, included the Socialisation and Communication subscales of the Vineland Adaptive Behaviour Scales (Sparrow, Balla, Cicchetti, Harrison, & Doll, 1984). Hierarchical multiple regression revealed that, after controlling for IQ and autism spectrum disorder features, social adaptive functioning only correlated with emotion recognition performance in extreme cases such as those observed in individuals on the autism spectrum and not those with nonspecific ID. García-Villamisar and colleagues (2010) speculate that these results may be the result of a 'threshold effect', where emotion processing deficits only disrupt social adaptive functioning beyond a certain level of severity.

The remaining five identified studies employed correlational analysis, with only two demonstrating a significant relationship between social functioning and facial emotion recognition performance (Rojahn et al., 2006; Rojahn & Warren, 1997). Rojahn & Warren (1997) observed a moderate correlation between emotion recognition performance on a facial emotion matching task, and measures of empathy and social responsiveness within the Social Performance Survey Schedule (SPSS Appropriate Social Skills subscale; Matson, Helsel, Bellack, & Senatore, 1983). In a more rigorous study, Rojahn and colleagues (Rojahn et al., 2006) presented four facial emotion recognition tasks which were differentially related to two measures of social adaptive functioning, the SPSS and the Vineland Adaptive Behaviour Scales Socialisation and Communication subdomains. Tasks included multichoice emotion picture to verbal label matching, emotion picture to emotion story matching, free emotion picture labeling, and emotion picture to picture matching. Only the latter two tasks produced a significant relationship with subscales measuring social adaptive functioning. The authors hypothesised that the observed inconsistencies may be due to poor ecological validity of the multichoice emotion picture to verbal label matching task and the emotion picture to emotion story matching tasks.

Little is known about the relationship between emotion recognition abilities and social adaptive functioning, both for individuals with ID and across the general population. Accurate interpretation of the discrepancies described in the few available studies is again compounded by issues of diagnostically diverse samples and task ecological validity. Most notably, there is potential that the inconsistencies observed across similar experimental tasks may be due to aetiological differences in the samples. Three of the six identified studies either did not explicitly define their sample or relied on samples of heterogeneous aetiology, with only one (García-Villamisar et al., 2010) explicitly meeting our criteria for nonspecific ID. In light of previously reviewed evidence of syndrome-specific patterns in emotion recognition, attempts to combine aetiological groups may have masked group specific strengths and weaknesses influencing the observed relationship with social adaptive functioning (Burack et al., 2016). This is of particular concern, should García-Villamisar et al., (2010) be correct in their hypothesis of a threshold effect.

Finally, we cannot discount the possibility that previously reviewed limitations in the ecological validity of common emotion recognition stimuli may too have masked a potential relationship between emotion recognition and social functioning. All six studies used black and white still photographs, meaning that stimuli were devoid of all colour and movement cues. If the ecological validity of emotion recognition tasks is limited, we cannot expect this to relate to an individual's day-to-day social functioning in a valid way. For example, positive findings may simply reflect more generalised cognitive capacities (e.g., information-processing speed and verbal abilities), while negative findings might be a result of failure to tap real emotion recognition capacities.

In summary, people with intellectual disability are among the most socially isolated in society (Wilson, Jaques, Johnson, & Brotherton, 2016). Fortunately, social adaptive functioning has been identified as a key predictor of social inclusion (Kozma et al., 2009). Should emotion recognition be identified as a key factor in social adaptive functioning this would present an important opportunity for targeted strengths-based intervention. However, before clinicians devote ongoing time and resource into such training, a more rigorous exploration of the link between facial emotion recognition and social functioning is required.

CHAPTER 5

OVERVIEW OF THE CURRENT STUDY

Emotion recognition ability has been studied extensively within the general population and to a fair degree in individuals with ID; yet, there has been a lack of interest in the development of a psychometrically sound measure capable of differentiating individual abilities to perform this important skill. A shortage of high quality measurement tools coupled with design limitations in existing ESH research has limited the legitimacy of findings regarding emotion recognition ability in individuals with ID, in addition to relevant outcomes such as social functioning.

Based on the previously reviewed limitations in existing research, the following requirements for the accurate assessment of the ESH can be derived: a) improved ecological validity through the application of dynamic colour representations of universal emotions, presented at a range of intensities; and b) improved experimental control through the inclusion of two participant groups (an experimental group with ID of homogenous aetiology and a typically developing mental age matched control group) combined with an experimental task of emotion processing and a control task of comparable abstraction and complexity. To the author's knowledge, there is no existing assessment measure appropriate for this task for which the psychometric properties have been established within an ID population.

To this end, the current study sought to develop and pilot an ecologically valid measure to assess facial emotion recognition abilities in people with ID, while attempting to reduce information-processing demands beyond those required to perceive the emotional content of stimuli. The decision to create a measure at the expense of the diversity afforded by a database was in the interests of developing a brief and uniform set of stimuli with potential for not only research but also clinical use. Items included in the new measure, herein referred to as the Kinetic Emotion Recognition Assessment (KERA), were selected based on consensus scoring, FACS criteria, and their scope to maximally capture individual differences and correlate meaningfully with pre-existing measures.

A heavy focus of the current study was also to assess how subtle methodological differences in the operationalisation of facial emotion may affect emotion recognition performance, and in turn provide insight into how we might reinterpret existing ESH literature. To this end, the KERA was applied in an investigation of the potential moderating effects of dynamic cues and emotion intensity, in addition to the assessment of the ESH. The overarching motivation for the revaluation and improved measurement of the ESH was in the interests of improving real-world outcomes associated with emotion recognition capacities. Accordingly, emotion recognition data were also interpreted in the context of two measures of social functioning to explore the link between social competence and emotion recognition ability. Table 1.1 provides a summary of the primary and secondary aims for this study.

CHAPTER 6

METHOD

The methodology is defined by two stages: Stage 1 refers to the development of the pilot measure; and Stage 2, the pilot testing and assessment of psychometric properties and subsequent item refinement of the KERA, followed by its application in a full assessment of the ESH and investigation of moderating factors. Figure 6.1 provides a summary of processes involved in KERA item development, validation, and application in hypothesis testing.

Ethical approval for the empirical studies reported in this thesis was obtained from Massey University Human Ethics Committee: Southern A, Application 13/67, and the Northern A Health and Disability Ethics Committee, Application 14/NTA/64.

Stage 1 - Development of the KERA item pool and selection of a core set

Development of the corpus. The following is a summary of the central features of the corpus, required to understand the selection and validation of the core item set used in the pilot measure.

Actors. The actors were 23 professional English speaking theatre and television actors, ten males and thirteen females. At the time of the recording, actors ranged between 16 to 69 years of age and included individuals of Caucasian, Indian, Asian, African, and Pacific Island (including Māori) descent. All actors received monetary compensation for their participation (25NZD per hour). To minimise the presence of extraneous visual stimuli in the corpus, actors were dressed in identical black shirts and all make-up, eyewear, and jewellery were removed. Men with beards were included only if the hair did not obscure the musculature of the face required to convey each of the six emotions. The inclusion of actors with facial hair was at the discretion of an independent Facial Actions Coding System (FACS; Ekman, Friesen, & Hager, 1978) certified consultant.

STAGE 1

Development of the KERA item pool and selection of a core set

↓

Development of the corpus

Facial behaviour from 23 actors depicting each of the six selected universal emotions was recorded. A total of 138 motion clips (23 per emotion category) were extracted from the footage to form the original corpus.

↓

Selection of a core item set

A core set of 24 items (4 per emotion category) were extracted from the corpus based on subjective ratings by nine independent observers followed by Facial Action Coding System (FACS) review.

STAGE 2

Pilot testing and item refinement of the KERA, and full assessment of the Emotion Specificity Hypothesis and investigation of moderating factors.

↓

Participant recruitment

Three participant groups each including 24 people took part in the study: An index group including community dwelling adults with nonspecific ID, and two control groups one matched for mental age and the other chronological age.

↓

Data collection

Experimental stimuli included seven discrete tasks (five emotion recognition tasks and two control tasks), in addition to psychological measures of social functioning. Several of the experimental measures were designed to serve a dual purpose, offering complementary but separate insights regarding each separate empirical investigation.

↓

Statistical Analysis

Part 1: Assessment of inter-rater item reliability estimates and subsequent item refinement of the KERA, followed by assessment of validity and appropriateness of KERA application in the target group (as defined by item difficulty and discrimination indices and group score distributions).

Part 2: Evaluation of the ESH and potential moderating factors (intensity and dynamic cues), followed by assessment of the relationship between social adaptive functioning and emotion recognition abilities.

Figure 6.1. Flow chart illustrating KERA item developmental, psychometric evaluation and application in the testing of the emotion specificity hypothesis.

Emotion elicitation procedure. Expression actors received a performance brief no less than 24 hours prior to recording. Instruction included detailed definitions of selected emotion categories (happiness, sadness, disgust, anger, fear, and surprise) as per the Oxford Dictionary of English – Third Edition (Stevenson, 2010) and temporal requirements for the emotional portrayals. Actors were required to evoke each of the six emotions separately, beginning with a neutral expression then developing the emotion to the highest intensity possible within the parameters of their personal range of emotion. Individual emotion displays were limited to 10 seconds. While displays of identical duration would have provided for superior experimental control, no further timing parameters were set, to ensure that the idiosyncratic temporal characteristics of each emotion were preserved. To avoid the production of stereotyped or exaggerated displays of emotion and achieve a balance of volitional and non-volitional control that may be expected in typical social interactions, actors were prompted to 'feel' each emotion allowing it to develop organically, and were not offered the opportunity to review the footage during the filming process. Beyond the requirements listed, actors were granted freedom to approach the task according to their own artistic process and were provided opportunity to repeat the task until satisfied that they had achieved an authentic performance. Invariably actors relied on personal recall and or mental imagery to provide emotive cues upon which to base their performance. Actors took an average of 1.5 hours to complete the task.

Apparatus. Facial behaviour was recorded using a Sony HDR-CX220E video camera. Actors were positioned 1.3 meters directly in front of the camera and filmed against a black studio backdrop set 3.75 meters from the camera and 2.45 meters from the subject. All ambient light sources were eliminated and a standard three-point lighting system was used to isolate the subject from the background and provide a flat lighting environment with as few cast shadows as possible. All three lights were positioned in front of the camera. The key light (principal light) was raised to a height of 1.8 metres and offset 45^{o} and 1.3 metres to the right of the camera. The fill light was raised to a height of 1.5 metres and offset 25^{o} and 1.0 metres to the left of the camera. The key and fill lights were filtered through a white diffusion shoot-through umbrella and reflector panel respectively. The back light was set 1.6 metres in front of the cameras (behind the subject), positioned 0.2 metres from the ground, and tilted upward at a 45^{o} angle. A set of barn doors (light modifiers) were fitted to the back light to reduce light scatter against the backdrop.

Image Processing. Post production video editing software Apple Final Cut Pro X (Apple Inc, 2013) was used to segment raw footage into brief video clips depicting discrete emotions. Clips were cut to include first a neutral expression displayed for approximately 0.5 seconds before the development of each emotion, and terminated at the apex of the emotion. The apex defined as the time point in the clip where the expression model showed the largest degree of muscle displacement from the neutral state. Clips were cropped to include the subject's head and neck only. Clips containing excessive head movement, blinking or indirect eye gaze were eliminated at the researcher's discretion. All clips where the subject's head moved into profile view were excluded. A total of 138 clips, one clip per expression model per emotion, were extracted to form the original corpus from which the core set were selected. Clip length ranged from between 2 and 10 seconds.

Selection of a core item set. A core set of items was extracted from the corpus based on subjective ratings by nine independent observers followed by independent FACS (Ekman, Friesen, & Hager, 2002) review.

Subjective ratings by independent observers. The full corpus was assessed by nine independent reviewers (four male and five female), including; three students enrolled in postgraduate clinical psychology training, two university staff members employed in the field of psychological and statistical research, and five additional reviewers with varied and unrelated expertise (Business Studies student, Landscape Architect, Community Sport Advisor, Artist and Administrator). Nine of the ten raters were of European descent and the remaining rater was of South African descent.

Individual clips from the original 138-clip corpus were embedded in FluidSurveys online survey software (SurveyMonkey, 2015). Raters were required to identify the primary emotion, and rate the naturalness of the performance. The emotion recognition task was presented in a forced-choice format and included the six universal emotions (happiness, sadness, disgust, anger, fear, and surprise) and a seventh category 'other'. The naturalness rating was presented as a sliding scale with the anchors 'completely contrived' (0%) and 'not at all contrived' (100%). Clip order was randomised for each respondent in addition to the location of response items for each trial.

Based on consensus scoring 50 clips were selected from the original corpus to undergo further review. Consensus scoring was deemed appropriate based on the underlying assumption of basic emotion theory regarding the evolutionary and social foundations of emotion and emotional expression (Ekman & Cordaro, 2011). In this regard, high levels of rater agreement were not only indicative of inter-rater reliability but may also be interpreted as evidence of content validity. Naturalness ratings were included in the interests of promoting face validity. Individuals who had ID were not employed as raters for this stage of analysis as emotion recognition capacities could not be assumed to be intact.

Group consensus was estimated using percent agreement and AC_1 (Gwet, 2008), a chance-corrected agreement coefficient designed for multi-rater nominal data. The interpretation of AC_1 is similar to that of the more popular generalized kappa statistic; however, unlike kappa, AC_1 avoids the first kappa paradox whereby high levels of observer agreement are met with low kappa values due to instability associated with trait prevalence (Feinstein & Cicchetti, 1990). In the current study, trait prevalence would refer to the number of raters selecting each of the six emotion categories. As high levels of consensus were desired, extremes in trait prevalence were anticipated and as such AC_1 was deemed the most appropriate test of inter-rater agreement. The computation of AC_1 throughout this study was conducted using the script file agree.coeff3.dist.r (Appendix B; Gwet, 2015) appropriate for RStudio (Version 0.99.902) Statistical computing software (RStudio, 2016).

Conservative bounds for the clip inclusion criteria were maintained in an effort to reduce the risk that high quality items might be excluded from further investigation as a function of the small number of raters. The minimum inclusion criteria based on inter-rater agreement was set at 70% correct and an AC_1 of at least 0.4 for the target emotion, in addition to a mean minimum naturalness score of 50%. An AC_1 of 0.4 aligns with the upper end of the 'fair' range if inter-rater reliability as per Altman's (1991) scale. Prioritisation of the selected 50 clips were based on negotiating clip naturalness rankings, and subject gender, age and ethnicity, in an effort to achieve the most diverse pool items possible.

Facial action coding system review. The Facial Action Coding System (FACS; Ekman, Friesen, & Hager, 2002) was applied to the remaining 50 clips by an accredited independent coder blind to the target expression.

The Facial Action Coding System is a manual observer-based system used to detect and taxonimise facial movements. The FACS system provides 46 core codes for all possible facial displays, which are referred to as Action Units (AU). For AUs that vary in intensity, a 5-point ordinal scale is used to measure the degree of muscle contraction. Intensities of AUs are annotated by appending letters A–E (A, trace; B, slight evidence; C, marked or pronounced; D, severe or extreme; and E, maximum evidence) to the Action Unit code. Additional AU codes are also available for the classification of head and eye movement, facial visibility and gross behaviour. These codes were not applicable to the current study. Appendix C presents a description of each of the main AUs and the underlying musculature.

FACS alone is a descriptive system and does not offer inferential emotion labels. To determine the content validity of the emotion portrayal in each clip, Emotion FACS (EMFACS; Friesen, & Ekman, 1983) was applied. Emotion FACS is a selective application of FACS scoring, in which coders only score behaviour that is likely to have emotional significance (i.e., AU code combinations associated with specific emotions). While FACS remains a subjective method, it is rigorously based on a description of facial movement and therefore may serve as a ground truth in expression recognition (Tian, Kanade, & Cohn, 2005). Table 6.1 presents the most common AU codes for each universal expression of emotion. Access to specific code combination rules for defining emotional expressions from FACS AUs is limited to accredited FACS coders, and as such has not reproduced in this thesis.

All clips that did not meet EMFACS criteria for the target emotion were excluded from further analysis. A total of 24 items were then selected for the pilot measure on the basis of demographic characteristics, the aim being to include as many high quality and diverse items as possible, while keeping the test at a length manageable for those with significant limitations in cognitive functioning. A total of 25% of the selected clips were cross validated by a second accredited FACS coder and perfect inter-rater agreement was observed. Reliability estimates for the final 24 clips, based on rating by the nine aforementioned independent raters, are presented in Appendix D.

Table 6.1

EMFACS AUs for Discrete Emotions

AU	Description	Emotion Labels					
		Surprise	Fear	Happiness	Sadness	Disgust	Anger
1	Inner Brow Raiser	✓	✓		✓		
2	Outer Brow Raiser	✓	✓				
4	Brow Lowerer		✓		✓		✓
5	Upper Lid Raiser	✓	✓				✓
6	Cheek Raiser			✓	✓		
7	Lid Tightener						✓
9	Nose Wrinkler					✓	
10	Upper Lip Raiser					✓	✓
11	Nasolabial Deepener				✓		
12	Lip Corner Puller			✓			
14	Dimpler						
15	Lip Corner Depressor				✓	✓	
16	Lower Lip Depressor					✓	
17	Chin Raiser					✓	
20	Lip Stretcher		✓				
22	Lip Funneller						✓
23	Lip Tightener						✓
24	Lip Pressor						✓
25	Lips part		✓				✓
26	Jaw Drop	✓	✓			✓	✓
27	Mouth Stretch	✓	✓				

Stage 2- Pilot testing and item refinement of the KERA and assessment of the emotion specificity hypothesis

Participants. A priori power analysis revealed that, an n of approximately 26 was required to obtain statistical power at the recommended .80 level with an alpha level of $p < .05$ (Cohen, 1992). Sample size was determined based on power to detect 'large' effect sizes in the data, as defined by Cohen's (1992) effect size conventions. Effect sizes were anticipated to be large in magnitude due to the high degree of correlation expected between the KERA and convergent validity measures, in addition to the mean between-group comparison effect sizes observed in the Scotland et al. (2015) review comparing emotion recognition abilities of individuals with ID and typically developing controls.

Participants in the current study included 24 adults with ID in addition to two control groups, one chronologically age matched, the other mental age matched. For the sake of parsimony, participants with ID will hereafter be referred to as the index group. Participants were also matched for gender and ethnicity, resulting in three groups each consisting of 9 males and 15 females, whereby 23 of 24 participants were of European descent and the final member of Pacific Island descent.

It is of note that the number of participants involved in the study fell just short of the quantity identified in the power analysis. Participant recruitment efforts were limited due to ethical obligations concerning consent processes, whereby participants were only to be approached through the identified support agency. While additional participants were available through the designated agency, this would have led to the inclusion of individuals with a more diverse range of developmental and mental health difficulties, and as such a smaller sample was tolerated in an effort to maintain tighter experimental control. The total number of participants with ID included in the study remains comparable to key related studies in the existing literature (e.g., Rojahn & Rabold, 1995; Owen, Browning, & Jones, 2001; and Scotland et al., 2016).

Participant eligibility and exclusion criteria.

Index group – Participants with Intellectual Disability. The current study is defined by the definition of intellectual disability assigned by the American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders – Fifth edition (American Psychiatric Association, 2013). Eligibility for participation in the index group required a pre-existing diagnosis of Intellectual Disability of mild severity and unknown aetiology (i.e. nonspecific ID), as indicated by specialist assessment and or diagnostic report. Individuals with ID of mild severity were the target population for the simple reason that individuals with nonspecific ID most often present with mild severity (Weis, 2013). Criteria included limitations in intellectual functioning, ongoing adaptive functioning deficits in various skill areas, and evidence that disability was present during the developmental period (i.e., before 18 years). Due to issues of confidentiality, the recruiting agency completed the screening phase for all potential participants, and specific diagnostic details, such as IQ (intelligence quotient) scores and diagnostic severity specifiers (based on adaptive functioning deficits) were not released to the researcher. Recruitment commenced in 2014. As such, participants would have been largely diagnosed per DSM-4 criteria whereby severity specifiers were assigned according to full scale intelligence quotient (IQ) score groupings. Thus it was inferred from the participant's attached 'mild' status that IQ scores ranged between 50-55 and 70.

In the DSM-5 severity is defined by adaptive functioning rather than IQ score. In an effort to verify the severity of adaptive functioning deficits and support our claims that participant diagnoses were limited to that of mild severity as per DSM-5 criteria, each participant's assigned support worker completed select subscales of the Vineland Adaptive Behaviour Scales – Second Edition Parent/Caregiver Rating Form (VABS-2; Sparrow, Balla, & Cicchetti, 2005). The VABS-2 is designed to assess personal and social skills across the lifespan. The Parent/Caregiver rating form includes three main indices appropriate for measuring adult functioning; Communication, Daily Living Skills and Socialisation. An Adaptive Behaviour Composite score is provided when all three domains are administered. Raw scores are converted to standard scores (Mean = 100, *SD* = 15) for each domain and composite score. The test structure maps neatly onto the three broad domains of adaptive functioning recognised by the American Psychatric Assocation (Diagnostic and Statistical Manual of Mental Disorders-5th Edition, 2013); Conceptual, Practical, and Social, and is therefore an ideal measure to estimate adaptive functioning deficits in the index group.

The DSM-5 offers qualitative descriptors with which to establish adaptive functioning deficits in intellectual disability, though the manual does not offer a method of operationalising limitations numerically. For the purposes of the current study, the American Association of Intellectual and Developmental Disabilities (AAIDD) manual guidelines have been applied to establish cut-off criteria for adaptive functioning scores. To this end significant limitations in adaptive behaviour may be defined as "performance that is approximately two standard deviations below the mean of either (a) one of the following three types of adaptive behaviour: conceptual, social, or practical or (b) an overall score on a standardized measure of conceptual, social, and practical skills" (Schalock et al., 2010, p.43). To more precisely define the severity of deficits observed, classification criteria as per the VABS-2 manual were also applied. Adaptive levels corresponding to subdomain and composite standard scores are as follows: Low 20-70; Moderately Low 71-85; Adequate 86-114; Moderately high 115-129; and High 130-160. Further refined classification scores are also offered for the Low adaptive functioning range: Profound deficit 20-25>, Severe deficit 20-25 to 35-40; Moderate deficit 35-40 to 50-55; and Mild deficit 50-55 to approximately 70.

The index group mean subdomain and composite scores and corresponding percentile rank and classification descriptors are presented in Table 6.2. Scores across all subdomains and the composite fell within the Low range, and may be defined as mild to moderate deficits. The VABS-2 scores were well below two standard deviations from the mean indicating significant limitations in adaptive functioning, consistent with a diagnosis of ID as per the DSM-5. Further, scores bore striking resemblance to the mean scores of the mild ID norm referenced group (Communication: M = 41.4, *SD* = 19.9; Daily Living Skills: M= 56.8, *SD* = 9.3; Socialisation:

M= 56.4, *SD* = 13.7; and Adaptive Behaviour Composite: M = 49.9, *SD* = 12.1; [Sparrow et al., 2005]). Further details regarding the VABS-2 and its application in hypothesis testing are provided under Materials.

Table 6.2

Index Group Vineland Adaptive Behaviour Scales – Second Edition Subscale and Composite Scores

Domain	Mean standard score (Standard Deviation)	Percentile rank	Classification
Communication	43.53 (24.57)	<0.1	Low–Moderate deficit
Daily Living Skills	53.12 (20.89)	0.1	Low–Mild to moderate deficit
Socialisation	54.65 (21.78)	0.1	Low–Mild to moderate deficit
Adaptive Behaviour Composite	50.73 (18.12)	<0.1	Low–Mild to moderate deficit

Exclusion criteria for the index group included the presence of sensory impairment that may jeopardise performance on experimental tasks, physical disabilities, or a pre-existing diagnosis of personality disorder or autism spectrum disorder as defined by DSM-5 criteria. All exclusion criteria were assessed by the recruiting agency based on participant personal records, with the exception of autism spectrum disorder which was assessed using the recently published Diagnostic Behavioural Assessment for Autism Spectrum Disorder – Revised – English version (DiBAS-R; Sappok et al., 2014), which was completed by the participant's assigned support worker.

The DiBAS-R is a standardised DSM-5 based caregiver-report screening tool developed specifically for the assessment of autism spectrum disorder in adults with ID. Factor analysis yielded two consistent dimensions; social interaction and communication (SCI), and stereotypy, rigidity and sensory (SRS) abnormalities. The DiBAS-R has demonstrated adequate diagnostic validity, as reflected by an area under the curve of 0.89 and balanced sensitivity and specificity values of 81%. The DiBAS-R total score also compares well with related measures, and is significantly correlated with the Social Communication Questionnaire (r = 0.52), the Scale for Pervasive Developmental Disorders in Mentally Retarded Persons (r = 0.50), and the Autism Checklist (r = 0.59). Discriminant validity is evidenced by the absence of a correlation with the Modified Overt Aggression Scale (Sappok et al., 2014). Exclusion criteria or cut-off scores were applied as per author recommendations (Total Score ≥29). Due to the extensive high quality psychological assessment available to the recruiting agency for each of the participants, management staff selected not to be informed in instances where high levels of autistic traits were observed. However, a summary of VABS-2 scores were reported providing participants offered their consent.

Chronologically age matched controls. Chronologically age matched controls were included to provide a baseline for the emotion recognition capabilities of neurotypical adults. Adult controls were selected based on their age at the time of testing, and observed no more than a 5-year age difference with their matched pair. Mean age for participants with ID and chronologically age matched controls was 44 years 10 months (SD = 12 years, range = 24 years 6 months – 67 years 6 months) and 47 years 4 months (SD = 13 years 8 months, range = 22 years 3 months – 72 years 6 months) respectively.

The Raven's Standard Progressive Matrices (SPM; Raven, Raven, & Court, 2000) were administered to chronologically age matched controls as a cognitive control measure. The test consists of a series of multiple choice visual analogy problems. Each problem contains a matrix of geometric symbols with one symbol missing. The respondent is required to identify the missing symbol that completes the pattern. Performance is measured in terms of overall score, which may then be used as an index to determine an IQ score from normative test data.

The Raven's SPM was developed as a test of nonverbal abstract reasoning. It is widely regarded as the leading test of Spearman's g, the general factor underlying performance on all tests of cognitive functioning (Jensen, 2002) and observes high convergent validity with the

Wechsler Intelligence Scales (r =.74-.84; O'Leary, Rusch, & Guastello, 1991). The Raven's SPM has well established reliability. Internal consistency studies using either the split-half method corrected for length, or KR20 estimates, result in values ranging from .60 to .98 with a median of .90. The median test-retest value is approximately .82. Test-retest coefficients are provided for the following age groups: .88 (13 years plus), .93 (under 30 years), .88 (30-39 years), .87 (40-49 years), .83 (50 years and over) (Raven, & Court, 2000). The Raven's 2012 norms supplement (NCS Pearson, 2012) was used to infer full scale IQ from raw scores, and experimental data from participants found to score below the upper margin of borderline intellectual functioning (IQ≤84; Wechsler, 2008) were omitted from the study.

Additional exclusion criteria included the presence of sensory or physical disabilities that may jeopardise performance on experimental tasks, or a pre-existing diagnosis of personality disorder or autism spectrum disorder as defined by DSM-5 criteria. All additional exclusion criteria were assessed based on participant self-report, with the exception of autism spectrum disorder which was assessed based on the Adult Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001).

The AQ is a screening questionnaire designed to assess autism spectrum traits in the general population. The test consists of 50 statements, designed to assess five different areas of functioning: social skill, attention switching, attention to detail, communication and imagination. Statements are presented using a forced choice format and respondents are required to indicate their level of agreement; 'definitely agree', 'slightly agree', 'slightly disagree' or 'definitely disagree'. Responses are scored using a binary system, where the endorsement of a trait of autism is assigned a score of one and non-endorsement a score of zero, contributing to a maximum score of 50.

Baron-Cohen et al. (2001) reported a series of validity studies demonstrating the AQ to discriminate high-functioning autism cases from controls. These studies also found the total AQ score and its five subscale scores demonstrated good test-retest reliability following a two-week interval (r=.70), good internal consistency as measured by Cronbach's alpha (Communication =

.65; Social, = .77; Imagination = .65; Local Details = .63; Attention Switching = .67), and acceptably high sensitivity and specificity. A cut-off score of 32 correctly identified 76% of patients (sensitivity 0.77, specificity 0.74) when applied to a referred clinical sample.

Consistent with author recommendations a cut-off score of >32 was considered indicative of clinically significant levels of distress associated with traits of autism and applied as exclusion criteria. Participants were given the option of having their general practitioner (GP) notified should screening reveal high levels of autistic traits. It was stressed that the measures used were for research purposes only and did not offer the same information as a comprehensive clinical assessment. For this reason, the release of test information remained exclusive to the participant's GP, who arguably would be in an appropriate position to interpret specific test item responses in the context of the individual's personal history.

Mental age matched controls. Mental age matching was established based on the Peabody Picture Vocabulary Test – Fourth Edition Form A (PPVT-4; Dunn & Dunn, 2007). The PPVT-4 is primarily a measure of receptive vocabulary. The measure was selected to ensure that matched pairs were equally capable of managing task specific demands related to the comprehension of emotion labels and the verbal task instructions. Further, vocabulary knowledge correlates highly with performance on more general measures of intelligence and is commonly viewed as a proxy for IQ for typically developing populations (Marchman & Fernald, 2008). Nevertheless, additional control tasks were included in the design to control for task related demands unrelated to language abilities, and to compensate for the possibility that participants with ID may observe a more varied cognitive profile, whereby verbal abilities prove to be disparate with nonverbal abilities.

The PPVT-4 has been extensively tested with special populations. Specifically, individuals diagnosed with ID score on average almost two standard deviations (26 points) below the general population (M=100, SD = 15), consistent with the large body of research proposing a strong relationship between vocabulary knowledge and general cognitive ability. The PPVT-4 has established reliability; specifically, internal consistency (Cronbach's alpha α=.97; split-half coefficient after Spearman-Brown correction r = 0.94) and test-retest reliability (r =.93). Test construct validity can be inferred by comparing the average performance across age with the profile of growth and decline. The growth curve of average performance in the PPVT-4 normative sample follows the pattern typical of measures of crystallised abilities (such as vocabulary); specifically, median scores increased rapidly between the ages of 2 to 4 years, then increased steadily until age 30, where levels were maintained through to the early 60s before decreasing. The PPVT-4 also compares well with related tests including measures of expressive vocabulary (Expressive Vocabulary Test – Second Edition [Williams, 2007], r = .82) and oral language (Comprehensive Assessment of Spoken Language [Carrow-Woolfolk, 1995], r = 69; and the Clinical Evaluation of Language Fundamentals [Semel, Wiig, & Secord, 2003], r = .76) (Dunn & Dunn, 2007).

Standard PPVT-4 administration requires respondents to point to one of four pictures that best describe a series of target words presented by the examiner. Each form is divided into nineteen sets of twelve items arranged in increasing difficulty. A starting point is determined based on the respondent's chronological age or perceived mental age in cases where a vocabulary deficit may be expected. To determine the basal, all items in a set are administered, allowing for only a single error. The ceiling is established once the respondent makes eight or more errors in a single set. The raw score is determined by adding the number of correct responses between the basal and ceiling to the basal score.

Standardised administration was maintained for the index group. In the interest of resource and maintaining grounds to complete the screening phase based on opt-out consent practices, mental age matched controls were assessed in a class setting. The test was adapted to a pen and paper multi-choice format to allow for group administration. The basal was set two years below the youngest child in each testing group and the test was terminated when items corresponded to an age equivalent more than three years older than the eldest child. In instances where a basal or ceiling was not established, respondent data was destroyed and the individual was not invited to participate in the study. A total of 253 children between the ages of five and eleven years were screened. Selected mental age matched controls observed no more than a 12-month age difference with their matched pair as established by age equivalence scores on the PPVT-4. Mean mental age for participants with ID and chronologically age matched controls was 9 years 10 months (*SD* = 2 years, 4 months) and 10 years 3 months (*SD* = 2 years) respectively. The mean chronological age for mental age matched controls was 9 years and 6 months (SD = 1 year, 3 months).

The identified mental age matched controls were only enrolled in the study if they were considered to be of average or above average intelligence. School achievement served as a benchmark for cognitive control, whereby potential participants were required to have met National Standards across all core academic subjects (reading, writing and mathematics). Additional exclusion criteria as assessed by parent report, included the presence of sensory or physical disabilities that may jeopardise performance on experimental tasks, or autism spectrum disorder.

Autism spectrum disorder was formally screened using the Autism Spectrum Quotient – Children's Version (AQ – Child; Auyeung, Baron-Cohen, Wheelwright, & Allison, 2008). The AQ-Child is a 50 item parent-report questionnaire that aims to quantify autism traits in children 4–11 years old. The scale adopts a similar format to the AQ-Adult, with the exception of specific item content and the scoring scheme. Similarly to the AQ-Adult, parents rate each statement to indicate their level of agreement 'definitely agree', 'slightly agree', 'slightly disagree' or 'definitely disagree', but in place of a binary scoring system each statement represented a score of 0-4 with higher scores indicating higher endorsement of autism traits. Possible scores range from 0 to a maximum of 150 suggesting full endorsement of all autism traits.

The AQ-Child has established reliability, with good test-retest reliability across a twelve-week period (r= 0.85), and excellent internal consistency (Cronbach's alpha coefficient = .97). Receiver-operating-characteristic analyses demonstrated that using a cut-off score of 76, the AQ-Child has high sensitivity (95%) and specificity (95%) (Auyeung, Baron-Cohen, Wheelwright, &

Allison, 2008). Consistent with author recommendations a cut-off score of >76 was considered indicative of clinically significant levels of distress associated with autistic traits and applied as exclusion criteria. Again, participant caregivers were given the option of having their child's general practitioner (GP) notified should screening reveal high levels of autism traits. Assessment and screening tools used in course of recruitment are summarised in Table 6.3.

Table 6.3

Participant Intake Assessment and Screening Measures Organised by Experimental Group

Assessment Domain	Group		
	Index Group	Mental age matched controls	Chronologically age matched controls
Cognitive Screen	Diagnostic or specialist report	National Standards	Raven's Standard Progressive Matrices
Adaptive Functioning	VABS-2	NA	NA
Mental Age Matching	PPVT-4	PPVT-4	NA
Autism Screen	DIBAS-R	ASQ-Child	ASQ-Adult

Note. NA = Not Applicable

Participant recruitment. Participants with ID were recruited from New Zealand's leading provider of services for people with intellectual disabilities. Participants were drawn from three separate branches offering supported independent living services, who responded to written invitation (Appendix E-1). Potential participants were first approached by an assigned support worker whom they had known for a minimum of six months (M = 3.5 years, SD = 2.75 years). The support worker explained the study guided by a pictorial information sheet and consent form (Appendix E-2). Interested parties, accompanied by their support worker, then met with the researcher who repeated this process and obtained written consent. All participants were deemed fit to provide informed consent which was seconded by their respective Area Managers (Appendix E-3).

Chronologically age matched controls were recruited via advertisements in the public domain. These included academic study management websites (www.callforparticipants.com, www.researchstudies.co.nz), university websites (www.massey.ac.nz), local newspapers (North Shore Times) and free weekly publications (Coffee News). The information sheet and consent form for chronologically age matched control participants are included in Appendices E-4 and E-5.

Participants for the mental age matched control group were recruited from three separate primary schools (decile rating = 7-10), who responded to a written invitation (Appendix E-6) distributed to 56 schools city wide. Parents of students enrolled in mainstream classes expected to contain children of appropriate mental age, received an information sheet (Appendix E-7) detailing the nature of the study with emphasis on the match screening phase. In accordance with the Principals' recommendation, the screening phase of the study operated on an opt-out clause, whereby parents were only required to respond to the letter if they wished to withdraw consent. The information sheet and consent form for mental age matched control participants and their parents, for the second part or experimental phase of the study, are included in Appendices E-8 to E-10.

Materials. The following is a summary of experimental tasks and psychological measures relevant to the validation of the KERA and assessment of the ESH and potential emotion recognition moderating factors. Several of the experimental tasks were designed to serve a dual purpose, offering complementary but separate insights regarding each separate empirical investigation. Table 6.4 provides a summary of all measures and assessment tools used to evaluate dependent variable outcomes. The assessment measures described in Table 6.4 were included for the purposes of hypothesis testing; however, many served a dual purpose and were also applied in the assessment of participant eligibility. The application of these measures in establishing inclusion and exclusion criteria has been previously defined in Table 6.3.

Table 6.4

Application of Experimental Measures Organised by Assessment Domain

Assessment domain	Assessment measures
KERA validity	
Convergent validity	KERA vs PoFA, NimStim and DaFEx (I)(MA)(CA)
Predictive validity	KERA vs DiBAS (I), AQ-Child (MA), and AQ-Adult (CA)
Emotion Specificity Hypothesis	
Emotion recognition measures	KERA, PoFA, NimStim and DaFEx (I)(MA)(CA)
Control measures	Age discrimination, Colour discrimination (I)(MA)(CA)
Effect of dynamic cues of emotion recognition performance	KERA vs KERA-Static (I)(MA)(CA)
Effect of emotion intensity on emotion recognition performance	KERA and KERA-Static (I)(MA)(CA)
Relationship between social functioning and emotion recognition performance	
Emotion recognition measures	KERA, PoFA NimStim and DaFEx (I)(MA)
Social functioning measures	VABS-2 Socialisation Index (I)(MA), SPSS (I) and DiBAS-R (I)

Note. I=Index group, MA=Mental age matched controls, and CA=Chronologically age matched controls; KERA=Kinetic Emotion Recognition Assessment, PoFA=Pictures of Facial Affect, DaFEx=Database of Facial Expressions, NimStim=NimStim Face Stimulus Set; DiBAS-R=Diagnostic Behavioural Assessment for Autism Spectrum Disorder - Revised (English version); AQ-Child=Autism Spectrum Quotient - Children's Version; AQ=Adult Autism Spectrum Quotient; SPSS=Social Performance Survey Schedule; VABS-2 = Vineland Adaptive Behaivour Scales – Second Edition.

Experimental tasks. Experimental stimuli included seven discrete tasks, five emotion recognition tasks and two controls tasks. Stimuli were embedded within Qualtrics survey software (Qualtrics Software [Version 37892], 2013) to form a multi-choice digitized test designed for use with a touch screen interface. The objective was to maintain a simple interface with minimal visual distraction whereby the only visual targets were the test stimuli set against a black background and the response options. Stimuli were preceded by a white blinking 'X' which remained on the screen for 1.5 seconds and served to draw the respondent's visual attention. Response options were presented as 'buttons' comprised of plain text within a white rectangle and positioned directly below the presented stimuli. The location of the response buttons remained fixed across all trials to reduce unnecessary cognitive load for the index group. Appendix F provides an exemplar of the experimental task interface.

Emotion recognition tasks. Five emotion recognition tasks were included in the test battery. Stimuli included the original 24 items of the KERA in both a static (KERA-Static) and dynamic form, and 18 items drawn from each of the Pictures of Facial Affect dataset (PoFA; Ekman & Friesen, 1976), the NimStim Face Stimulus Set (Tottenham et al., 2009) and the Database of Facial Expressions (DaFEx; Battocchi et al., 2005). Stimuli sets rather than standardized test of emotion were selected, due to a limited availability of a valid and reliable measure of emotion decoding that exclusively relied on facial cues and used simple response formats. Further, the wide variety of emotion samples available in each of the selected stimuli sets supplemented experimental control, permitting for some matching across emotion intensity, race, and still versus dynamic stimuli, while allowing for the response interface to remain consistent.

Dynamic and still forms of the KERA were presented as separate tasks. The static measure took an identical form to the original video-based measure with the exception that only the last frame of each clip was presented as test stimuli. Static stimuli remained on the screen for the same duration as the original clip from which it was derived. The two versions of the tasks served as a means of exploring potential moderating effects of dynamic cues on the emotion recognition, while controlling for factors such as gender, ethnicity, emotion intensity and additional idiosyncratic features of the expression model.

The PoFA, NimStim and DaFEx stimuli were primarily included to provide a benchmark for the assessment of convergent validity of the KERA, in addition to offering additional assessment of emotion recognition ability with which to investigate the ESH. In the interests of brevity, only three emotion representations per emotion category were selected from each of the PoFA, NimStim and DaFEx, in contrast with the four emotion representations included in the KERA. To reduce the impact of possible demographic confounds when comparing performance on the KERA and the convergent validity tasks, three of four items from each emotion category of the KERA were randomly selected and effort was made to match for demographic characteristics (race, age and gender) of actors and item intensity across the stimuli sets. In instances where this information was not available inferences were made by the primary researcher. Clip length was also controlled for using this method, with the exception of the DaFEx where clip length was predetermined due to the stimuli taking the form of short video clips. The selected stimuli from the PoFA, NimStim and DaFEx are listed in Appendix G.

The PoFA database is one of the most widely used databases in emotion research, and is heavily utilised within the ID literature (for a review of ESH studies see Moore [2001] and Scotland, McKenzie, Cossar, Murray, & Michie [2016]). The database includes 110 black and white photographs of Caucasian actors. All the emotions were simulated upon request using FACS based instruction, with the exception of happiness which was derived from spontaneous expressions. All expression samples were FACS verified to confirm that they included the correct configuration of AUs. While the PoFA is not formally a standardised test it has been used as an individual differences measure of emotional sensitivity in a large number of studies, including developmental (Cheal & Rutherford, 2011; Pollak & Kistler, 2002), neuroimaging (Howard et al., 2000), and behavioural research (Dimberg, Thunberg, & Elmehed, 2000). Generally, the psychometric properties of the PoFA are largely assumed, as its construction was based on the

FACS system. The PoFA also demonstrates high inter-rater agreement. Ekman (1976) reported that 90% of the pictures were correctly rated by more than 80% of normal functioning adults, and the mean accuracy was 88%. Mean inter-rater agreement for specific items selected for the current study was 93.2% (*SD* = 7.4).

The NimStim Face Stimulus set includes 646 colour photographs posed by Americans of European, Latino, African and Asian descent. As with the PoFA stimulus set, actors were asked to adjust specific muscles to produce the desired emotions to a high intensity. Each expression in the database is offered in an open and closed mouth form, with the exception of surprise for which all photos included an open mouth. Unlike the PoFA, the NimStim is not FACS verified. However, the set is arguably more contemporary in appearance, is available in colour and offers a selection of stimuli derived from non-Caucasian expression models. Empirical support for the validity and reliability of this set is evidenced by the accurate identification of expressions in the founding study (mean proportion correct = 0.81, *SD* = 0.19, mean kappa across stimuli = 0.79, *SD* = 0.17) and high intra-participant agreement across two testing sessions (*M*= 0.84, *SD* = 0.08) (Tottenham et al., 2009). Mean inter-rater agreement for items included in the current study was 91% (*SD* = 7.4).

The DaFEx corpus is comprised of 1008 short videos (4-27 seconds) depicting emotions at three different intensities (low, medium and high) expressed by eight Italian actors. Intensity ratings were based on the subjective opinion of the expression model and only high intensity stimuli were selected for the current study. The corpus includes two conditions: an utterance condition where actors performed emotions while uttering a phonetically rich and visemically balanced sentence, and a no-utterance condition where actors performed emotions in the absence of any dialogue. The emotion representations began and ended with the actor showing a neutral expression. Stimuli for the current study were borrowed from the no-utterance condition. Inter-rater agreement for the no-utterance condition, as measured by 80 untrained respondents, was observed at 75% across the six universal emotions. Scores for discrete emotions ranged from 69% for fear, to 78% for surprise (Battocchi et al., 2005). The standard deviations for reported means and or chance corrected agreement statistics were not provided in the initial validation study, and inter-rater reliability scores for individual clips were not made available to the researcher when the stimuli were received. The DaFEx was included in the study despite the limited availability of item level data, for the simple reason that the DaFEx stimuli offered the novel edition of video-based stimuli, making it an appropriate tool to corroborate any findings regarding the moderating effects of dynamic cues as they pertained to the ESH.

Control tasks. Two control tasks were included in the test battery; a) an age discrimination task and b) a colour discrimination task. Control tasks were parallel forms of the emotion recognition tasks devoid of emotional content, administered to control for the cognitive demands of experimental tasks. Control task stimuli were organised to maintain the mean stimuli duration observed in the emotion recognition tasks. Each of the control tasks included 12 items in contrast with the 24 included in the emotion recognition tasks. To compensate for the reduced number of items, the emotion recognition task items were ordered according to their clip length and the mean clip length for each successive pair then applied as the timing criteria for control task items.

a) Age discrimination task: In the age discrimination task, respondents were required to identify the chronological age of a human target. The purpose of the age discrimination task was to determine whether any difficulties observed in facial emotion identification were due to domain specific emotion processing deficits or a more general deficit in the processing of human faces. Chronological age, like emotion, is one of the few physical human attributes where small changes in form result in a significant change in meaning (Hobson, 1991). The inclusion of a face based age discrimination task allowed inferences to be made regarding the specificity of any potentially observed deficits in emotion processing, by controlling for the visual complexity of facial stimuli.

Stimuli were taken from the Centre for Vital Longevity Face Database (Minear & Park, 2004), and included static colourful images of people from various age groups displaying a neutral expression. Images were altered to include a black background consistent with the emotion recognition task stimuli. The age discrimination task was presented in an identical format to the emotion processing tasks with the exception that only three age response categories were included; Young Adult, Middle Aged Adult and Old Adult. The use of six response categories, as applied in the emotion recognition tasks, was not included, since age as a construct does not lend itself to be divided into six discrete categories. Such a response format would call for the use of several numerical age ranges (e.g., 40-50 years old) which is arguably more difficult to comprehend than three discrete emotion categories. Four representations for each age group were included in the task.

b) Colour discrimination task: The colour discrimination task was administered to compensate for the limited number of response categories offered by the age discrimination task, and to control for participant abilities to hold dynamic visual information in mind, access meaning across modalities (visual and verbal), and select a response from five distracter categories. The task included a series of two-dimensional computer animations developed using graphics editing software Adobe Photoshop (Version CS5). Images depicted the transitioning of one geometric shape (oval, square, triangle, hexagon, circle or rectangle) into another. The effect was achieved by reducing the opacity of the first shape while increasing the opacity of the second in successive frames. Shape colour remained constant across each trial and hues included red, yellow, purple, green, orange and blue. Two trials were presented for each colour including one dark and one light shade. All stimuli were set against a black background. Appendix H includes an example of the animation at five different time points. The colour discrimination task was presented in an identical format to the emotion processing tasks, whereby the response options were the six available colour options. Respondents were required to identify which of six colour response categories best described the stimuli.

Psychological measures. Several of the instruments applied in hypothesis testing served a dual purpose and have been previously described in the course of participant recruitment. Select measures that require further elaboration are described here. A complete list of assessment measures included for the purposes of participant recruitment and hypothesis testing have been previously defined in Table 6.3 and 6.4 respectively. To explore the relationship between emotion recognition abilities and social performance, general measures of social adaptive functioning (VABS-2, Social Performance Survey Schedule [SPSS], and the DiBAS - Social Interaction and Communication subscale) were applied. In addition, autism spectrum disorder screening measures (DiBAS-R, AQ-Child, AQ) were included to assess the predictive validity of the KERA.

Assessment of social functioning was limited to the index group and mental-age matched controls, due to logistical issues associated with recruiting a third party informant to report on chronologically age matched controls. Further, it was anticipated that chronologically-age matched controls would observe ceiling effects on experimental tasks, limiting statistical analysis of the relationship between social functioning and emotion recognition abilities.

To assess for social functioning, select subscales of the Vineland Adaptive Behaviour Scales – Second Edition (VABS-2; Sparrow et al., 2005) parent/caregiver rating form were completed for both the index group and mental age-matched controls. In addition to the VABS-II, the index group was also assessed on the Social Performance Survey Schedule (SPSS; Matson, Helsel, Bellack, & Senatore, 1983), and scores on the Social Interaction and Communication subscale of the DiBAS were also considered. The priority was to explore the social implications of potential emotion recognition deficits. Accordingly, the assessment of social functioning in the index group was somewhat more rigorous than that afforded the mental-age matched group, who were unlikely to present with emotion recognition difficulties. Measures were completed by the assigned support worker for index group participants and by a legal parent or guardian for mental aged matched controls. Support worker details are the same as those described under Participants.

Vineland Adaptive Behaviour Scales – Second Edition. The VABS-2 is designed to assess personal and social skills across the lifespan. The VABS-2 is a widely used informant based measure of adaptive functioning (García-Villamisar et al., 2010) and observes robust psychometric properties when used with both typically developing individuals (Sparrow et al., 2005) and those with ID (de Bildt, Kraijer, Sytema, & Minderaa, 2005). Supplementary norms are available for every level of intellectual disability across the lifespan. The parent/caregiver rating form includes five indexes: Communication, Daily Living Skills, Socialisation, Motor Skills, and Maladaptive Behaviour. The Socialisation index falls under the rubric of social adjustment and was selected for this study. The index is comprised of three subscales: Interpersonal Relationships (How the individual interacts with others); Play and Leisure Time (How the individual plays and uses leisure time); and Coping skills (How the individual demonstrates responsibility and sensitivity to others). VABS items have weighted ordinal rating options (no, never; sometimes or partially; yes, usually; don't know; or no opportunity [to observe]) that follow a developmental sequence. Higher scores are associated with superior adaptive functioning.

The Socialisation subdomain has demonstrated high levels of internal consistency (Split-Half Coefficient using Nunnally's formula for ages 18-90 = .84-89) and inter-rater reliability (intraclass correlation coefficient for ages 18-90 = 0.66-0.69) (Sparrow et al., 2005). Validity evidence is provided by the demonstration of a systematic relationship between test scores and diagnostic severity criteria for ID. For every level of increasing diagnostic severity in a clinical sample, the VABS-II demonstrates increasing clinical deficits across all domain scores and the composite, all of which exceed two standard deviations below the mean score of a nonclinical reference group. When comparing mild and moderate ID, both groups reflect theoretical expectation, displaying a flat pattern of deficits across domain scores (Sparrow et al., 2005).

Social Performance Survey Schedule. The SPSS by Matson and colleagues (1983) is a revision of the 100 item SPSS designed by Lowe & Cautela (1978) to assess the positive and negative social behaviours of adults of normal intellectual functioning. The informant rating survey was adapted for use with adults with mild-moderate intellectual disability and includes 57 items designed to assess strengths and weaknesses in higher order social skills. These items were retained based on Pearson product-moment correlations of .30 or greater with the total score of the original measure.

Typically, the SPSS is administered as a structured interview. However, for our purposes the measure was completed independently by respondents. The researcher remained in close vicinity to offer clarification when required. Responses on the SPSS are rated on a five-point Likert scale (0=Not at all, 1=A little, 2=A fair amount, 3 = Much, 4= Very much) and are organised into four factors which are further grouped into two subscales; Maladaptive Behaviour subscales (Sociopathic Behaviour and Inappropriate Assertion) and Prosocial Behaviour subscales (Communication and Appropriate Social Skills). High scores correspond to higher prevalence of behaviours associated with each of the four domains. The Maladaptive Behaviour subscales do not assess social adaptive functioning however have been included in the study for interest, due to the integrated format of the measure meaning that it was best presented in its entirety.

The SPSS has good inter-rater reliability (positive items .71, negative items .69; Matson et al., 1983) and high levels of internal consistency (Cronbach's alpha = .88, Guttman Split-Half Coefficient = .87; Matson, 2009). Unfortunately, despite being one of the only brief measures of social performance developed for adults with mild and moderate ID, there is no available normative data or research supporting the validity of the revised version of the scale. Consistent with the basic premise of this thesis, it is again evidenced that well validated brief measures properly adapted or designed for use with ID remain limited. For lack of an alternative the SPSS was employed to qualify results established using the Vineland-II.

Design and experimental procedures.

Design. The study employed a matched pairs quasi-experimental design, whereby the application of emotion recognition and control tasks was by within-subjects design, and measures of social adaptive function both within- and between-subjects. Assessment of social functioning was limited to the index group and mental age matched controls. Here the VABS-2 was used to assess both the index group and mental age matched controls, while the SPSS was applicable only to the index group. Application of measures assessing traits of autism used to evaluate the predictive validity of the KERA, applied a between-subjects design. The independent variable in this study was group membership, and dependent variables included performance scores on measures of emotion recognition and the control tasks, and informant ratings on measures of social adaptive functioning and autism traits. To mitigate order effects on experiment tasks, task order and the order of individual test stimuli were randomised but remained consistent across each matched pair.

Experiment Procedures. Data collection was completed by the primary researcher and a research assistant. The primary researcher had previously held employment working with individuals with ID in both an educational and therapeutic setting. At the time of data collection, the research assistant was completing their final year of the Masters of Education Psychology qualification. The research assistant who already held a Masters qualification in primary school teaching and had extensive experience as a qualified primary school teacher, also served as a consulting party regarding the procedures surrounding student interactions with mental age matched controls. Participant consent and testing was completed in a single sitting and in cases where inclusion criteria were not met, participant data was anonymously removed after the fact to avoid distress. In addition to third party consent procedures established prior to the testing session, all participants provided written consent immediately prior to testing. Table 6.5 summarises data collection procedures and associated use of assessment measures, organised by experimental group.

Table 6.5

Data Collection Procedures and Application of Assessment Measures Organised by Experimental Group

Procedures	Index Group	Mental age matched controls	Chronologically age matched controls
Recruitment	Participants were recruited from a specialised ID service provider.	Groups of potential participants from three local primary schools were assessed for mental age based on performance on the PPVT-4. Individuals appropriately matched for mental age, gender and race were invited to participate.	Participants recruited from advertisements in the public domain.
Testing session	Participants completed an assessment of mental age (PPVT-4) and the experimental tasks. The assigned support worker completed social functioning (VABS-2 and SPSS) and autism spectrum disorder screening measures (DiBAS-R).	Participants completed the experimental tasks.	Participants completed all seven experimental tasks and screening measures (cognitive screen–Raven's SPM; autism spectrum disorder screen - AQ).
Post testing session	Data was excluded in cases where exclusion criteria were met, and third parties informed where necessary.	Measures of social functioning (VABS-2) and screening measures (autism spectrum disorder; AQ-Child) were completed by parent or guardian. Data was excluded in cases where exclusion criteria were met (n=1), and third parties informed where necessary.	Data excluded in cases where exclusion criteria were met (n=1), and third parties informed where necessary.

Note. PPVT-4=Peabody Picture Vocabulary Test-Fourth Edition; VABS-2=Vineland Adaptive Behaviour Scales–Second Edition; SPSS=Social Performance Survey Schedule; DiBAS-R=Diagnostic Behavioural Assessment for Autism Spectrum Disorder-Revised (English version); AQ-Child=Autism Spectrum Quotient- Children's Version; AQ=Adult Autism Spectrum Quotient.

Psychological screens and measures were delivered via post to the parents or guardians of mental age matched controls. Materials sent by post were anonymised to avoid potential breach of confidentiality and returned through the participating primary schools. Chronologically age matched controls completed psychological measures immediately following the testing session, while index group support workers completed measures at the time of testing. All parties involved were provided with the coordinating researchers' contact details and were encouraged to seek clarification if required.

Testing sessions were completed in a quiet room in a one-to-one setting. The index group were provided with the option of having testing sessions completed in their homes or at the offices of the recruiting agency. Mental age matched controls were tested at school, and chronologically matched controls at the supervising tertiary institution. The index group were assessed on the Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4), used for mental-age matching, at the outset of the formal testing session. Recall that the mental age matched controls completed the PPVT-4 assessment on a separate occasion being that performance on this test determined eligibility to participate in the remainder of the study.

Following on from the consent process (and completion of the PPVT-4 for index group participants), participants completed the seven computerised experimental tasks. In preparation, a criterion task was first completed to accustom participants to the task of labelling facial emotion and to assess emotion expression familiarity. Participants were required to enact the facial expressions consistent with the six universal emotions, and then to label the examiner's facial expressions depicting these same emotions. For failed emotions, participants were provided with the emotion definition as per the Oxford Dictionary of English – Third Edition (Stevenson, 2010), whereby the researcher was afforded creative licence to amend the explanation to assist comprehension. Common emotion-eliciting scenarios were also discussed and participants were encouraged to develop personal examples to foster understanding. The scenarios, adopted directly from earlier work by Rosenberg and Ekman (1995), are reproduced in Table 6.6. Participants only continued on to complete the experimental tasks if they were capable of identifying, producing or describing each of the six emotions. Data belonging to one participant from the index group was omitted from the study based on these criteria.

Table 6.6

Emotion Eliciting Scenarios

Story name	Target emotion	Scenario
Friends	Happiness	The person sees many close friends at a party.
Child	Sadness	The person remembers the time last year when a young child died of terminal illness.
Dog	Disgust	The person steps in dog faeces, reaches down to wipe it off, and faeces gets on the person's hand.
Post Office	Anger	The person is waiting in line at the post office for a very long time. The person finally reaches the window, when the clerk announces that there is time for only one more customer. The person is then pushed aside when someone cuts in front to take the person's turn.
Brakes	Fear	The person has realised that the brakes don't work while driving down a steep hill. The car approaches the end of the road, which is a cliff with no barrier. The person tries to brake and veers out of control.
Tall	Surprise	The person is sitting next to someone who suddenly stands up and is much taller than the person had expected.

Following on from the criterion task participants completed the experimental tasks. These were presented on a touch screen tablet computer (iPad Third Generation [Operating System iOS 5.1]). Owing to the identical format of each discrete task, as a rule practice trials only preceded the first emotion recognition task presented and the two control tasks. Additional practice trials were included at the researcher's discretion. The researcher only continued with test trials once confident that the participant was able to participate in a discrimination task by issuing a pointing response or verbally responding. Participants were afforded the opportunity to replay test stimuli to reduce the potential confounding effects of memory deficits and fatigue. The following instructions were given prior to each trial or until the participant became familiar with the response format:

> "I am going to show you some videos (or pictures), one at a time. The videos show people feeling different emotions"
>
> "I would like you tell me if the person in the video is feeling happiness, sadness, fear, surpise, anger, or disgust". [Researcher points to the appropriate emotion labels on the screen]
>
> "You may tell me or point to word on the screen if you like".

This script was adapted for the age discrimination and colour discrimination control tasks. Prior to the colour discrimination task, participants were provided with a colour swatch depicting the six colours included in the test and completed a simple labeling task to ensure familiarity. Participants displayed a unanimous understanding across all colour categories. Similarly, before completing the age discrimination task, particiapants were provided with a second swatch listing the age response categories and corresponding numerical age ranges (Young adult 18-30 years; Middle Aged Adult; 40-60 years; Old Adult 70≤years). To assess comprehension of task requirements, participants were prompted to provide examples of people from their own lives (verified by the support worker) or well known public figures who fell in each of the age categories. Performance was noted, and all participant progressed with the task irrespective of their understanding of the different age categories. Participants in all groups were afforded breaks upon request. Individuals in the index group took an average of three breaks, while participants in the remaining two groups preferred to complete the tasks in a single sitting. All participants were presented with a retail voucher (10 NZD) offered as a token of appreciation for their participation.

Statistical procedures. Preliminary data screens included a random data integrity check and the assessment and management of missing data, outliers and data distribution. The subsequent statistical procedures may be considered in two parts. Part 1 refers to the psychometric evaluation of the KERA, and Part 2 an investigation into the nature of the emotion recognition abilities of people with ID and relevant outcomes in social functioning. Relevant thresholds for interpreting effect sizes are included in Appendix I. All data were analysed using the IBM Statistical Package for Social Sciences (SPSS) for Macintosh, Version 23.0, unless specified otherwise.

Part 1. Evaluation of the psychometric properties of the KERA. Part 1 procedures included the assessment of item reliability and subsequent item refinement, followed by evaluation of subscale and full scale reliability, item difficulty and discrimination, analysis of score distributions and finally validity assessment. Item and subscale reliability was assessed using identical parameters applied in the preliminary selection of clips from the original corpus. Reliability data were not derived from individuals with ID, as emotion recognition capacities could not be assumed to be intact. Instead, data were based on chronologically age matched adult controls. Again, the minimum inclusion criteria based on inter-rater agreement was set at 70% correct and an AC_1 of at least 0.4 for the target emotion. Qualitative descriptors for strength of agreement as defined by AC_1 estimates, were assigned to subscale and scale total scores, based on the Altman's scale depicted in Table 6.7 (Altman, 1991). The benchmark for agreement coefficients was calculated using Gwet's (2014) alternative benchmarking methods of cumulative membership probabilities.

The benchmarking procedure traditionally used by researchers is based upon a direct comparison between the calculated agreement coefficient and a number of possible benchmark scales proposing various qualitative descriptors with which to interpret strength of agreement. The classical approach tends to provide an inflated picture regarding the extent of agreement among raters, overlooking the adverse effects that a small number of raters or categories may have on the precision of estimated agreement coefficients (Gwet, 2014).

Table 6.7

Altman's Kappa Benchmark Scale

Kappa statistic	Strength of agreement
< .20	Poor
.21 - .40	Fair
.41 - .60	Moderate
.61 - .80	Good
.81 - 1.00	Very Good

In contrast, Gwet's alternative benchmarking model is probabilistic. That is, each benchmark level (i.e. qualitative descriptor range) of the selected scale is assigned a membership probability, representing the likelihood that the estimated agreement coefficient indeed falls into the associated range of values. The assigned probability takes into account the standard error associated with each individual agreement coefficient. The final level of agreement corresponds to the highest benchmark level associated with the smallest cumulative probability that exceeds a predetermined threshold (Gwet, 2014). In this case, 90% was selected as an appropriate cut-off point. Such a conservative approach has been applied in the current study as a means to eliminate doubt regarding potential adverse effects of the modest sample size, and ensure confidence in the minimum standard of reliability achieved by the scale.

Following reliability analysis, the discriminating power of test items and item difficulty were examined across the three experimental groups. Item discrimination values, as expressed by the point-biserial statistic (r_{pbis}), were calculated for each of the remaining items. The point-biserial correlation, which is the observed correlation between item performance and test performance, can range between -1 and 1; where more positive values indicate superior discrimination and values in excess of .20 are considered to be desirable (Hambleton & Dirir 2003). Ebel and Frisbie (1991) suggest that items below 0.19 may be considered poor items, 0.20

to 0.29 marginal items in need some revision, 0.30–0.39 good items with little or no revision required, and items with a discrimination index of 0.40< very good items. Group data was collapsed in the calculation of item discrimination values, in an effort to take advantage of the extended range of scores and to promote stability of the observed correlations through increased sample size.

In contrast, item difficulty is a characteristic of both the specific test item and the sample taking the test. Consequently, item difficulty indices, given as *p* values, were calculated separately for each of the three experimental groups. Item difficulty represents the proportion of participants who answered an item correctly, where lower *p* values denote more difficult items and higher values less difficult items (Wood, 1974). A *p* value of 0.0 or 1.0 indicates that the specific item is unlikely to contribute to measuring individual differences (Thorndike & Thorndike-Christ, 2009). Ideal item difficulty falls halfway between a perfect score and the proportion expected to answer correctly if pure guessing governed responses (Thompson & Levitov, 1985). The KERA offers six discrete response options. Accordingly, the ideal mean item difficulty to maximise score reliability is approximately: $\left(\frac{1}{6}+\left\{\left(1-\frac{1}{6}\right)\div 2\right\}\right)=.58$.

Building on individual item analysis, an examination of group score distributions was completed to assess the appropriateness of the KERA for use with adults with ID when compared with the two control groups. It was hoped that the index group would observe a wide spread of normally distributed scores with minimal floor or ceiling effects. Distribution symmetry and pointedness were calculated based on skewness and kurtosis statistics, and analysis of the overall spread of group score distributions was based on descriptive statistics (mean, standard deviation and range).

Assessment of convergent validity was then established by correlating total scores from the KERA with scores on the PoFA, NimStim and DaFEx. Correlational analysis was limited to full scale scores for two reasons. First, each emotion category does not represent an underlying trait or ability therefore one cannot assume a stable pattern of scores across emotion categories, and subscale score comparisons across the four stimuli sets would likely be confounded by a number of variables, including the presence or absence of dynamic cues, colour, expression model ethnicity, emotion intensity and overall item difficulty.

Finally, a simple linear regression was used to assess predictive validity, specifically how performance on the KERA predicted levels of traits associated with autism spectrum disorder. The focus here is on 'traits' associated with autism, being that autism spectrum disorder was part of the exclusion criteria for this study. A negative relationship was anticipated based on the well documented emotion recognition difficulties experienced by individuals with autism spectrum disorder (Uljarevic & Hamilton, 2013), and findings that the extent of autistism traits observed in typically developing individuals also predicts impairment in emotion identification (Poljac et al., 2012). Proportion of explained variance could not be predicted, due to a lack of available research employing correlational designs upon which to inform hypotheses.

Part 2. Investigating the nature of the emotion recognition abilities of people with ID. The objectives of Part 2 of the analysis were to: re-evaluate the applicability of the Emotion Specificity Hypothesis for individuals with Intellectual Disability; identify potential moderating effects of dynamic cues and emotion intensity on emotion recognition performance; and investigate the link between social adaptive functioning and emotion recognition abilities in adults with ID and typically developing children.

Full support for the Emotion Specificity Hypothesis required a significant performance discrepancy between the index and control groups on tasks of emotion recognition, coupled with comparable performance across the two control tasks. A series of one-way repeated measures Analyses of Variance (ANOVA) were conducted to determine whether there were statistically significant differences in emotion recognition performance between the index group and the two control groups. The within-subjects factor was group membership and the dependent variable, each of the three emotion recognition tasks. Repeated measures ANOVA were selected in place of independent ANOVA to remove variation due to individual differences, ensuring a more powerful and sensitive test. Each matched threesome was treated as a single sample member.

Repeated measures ANOVAs were also conducted to compare control task performance across the three groups. Post hoc analysis included a series of paired-samples t-tests.

An additional series of analysis of variance were completed to identify potential deficits specific to emotion category. Based on the distribution of the data nonparametric tests were selected. Subscale scores were collapsed across the KERA, PoFA, NimStim and DaFEx. The Friedman test was applied to determine whether any statistically significant differences existed between the three experimental groups across each of the six emotion categories. Post hoc pairwise comparisons were calculated using multiple Wilcoxon signed-rank tests.

The effect of dynamic cues on facial emotion recognition performance were examined by comparing group mean total performance scores on the KERA and the KERA-Static. Separate paired sample t-tests were used to identify within group differences highlighting the moderating effects of dynamic cues based on group membership. Then, to determine the effect of emotion intensity on facial emotion recognition performance, Pearson product-moment correlation coefficients observed between summed item total scores for the KERA and item intensity, were compared across each of the experimental groups.

Finally, correlational analysis was again applied to explore the link between social adaptive functioning and emotion recognition abilities in adults with ID relative to mental age matched controls. Total scores were combined across the KERA, PoFA NimStim and DaFEx to form an emotion recognition composite score which was then correlated against various measures of social functioning (Index group: VABS-2 Socialisation Index, SPSS and DiBAS-R Communication and Interaction subscale; Mental age matched controls: VABS-2 Socialisation Index). Correlations were also computed for the KERA and KERA-Static to explore additional hypotheses regarding whether the observed relationship may be moderated by dynamic cues as a result of enhanced ecological validity. Gender differences were not a central focus of this study; however, there is some evidence of gender differences across a number of emotional abilities and behavioural outcomes (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006). For example, emotion recognition accuracy has been found to be significantly related to social adjustment for girls, but not for the boys (Custrini & Feldman, 1989; Leppänen & Hietanen, 2001). Accordingly, correlational analyses were also conducted separately for males and females.

CHAPTER 7

RESULTS

Preliminary data screening

Data integrity check. Ten percent of the manually encoded data files were randomly selected, rescored where applicable, and checked against the corresponding digital records. All files were found to have been correctly scored and recorded, affording confidence in the accuracy of the remaining data.

Statistical outliers. Univariate and bivariate outliers were identified for all variables that followed an approximately normal distribution and were implicated in the use of parametric tests. Extreme univariate data points were identified with reference to the interquartile range (IQR) based on ranks. Outliers were considered to be mild with values between 1.5-3 times the IQR, and extremes cases were identified as values more than three times the IQR (Barbato, Barini, Genta, & Levi, 2011).

In the analysis of psychological measures, two outliers were detected within the AQ-Child and VABS-2 data derived from mental age matched controls. Both data points were removed from the data set. No univariate outliers were detected in the analysis of psychological measures for the index or chronologically age matched groups.

Analysis of outliers on experimental tasks was limited to full scale scores. Only two univariate outliers were detected on experimental tasks in the index group. Cases were mild and derived from the colour discrimination control task data. Due to overall high performance on the task, the two items fell only two points (out of a total possible score of 12) below the median. These cases are likely to represent 'real' results rather than anomalous data and were therefore considered valuable in regard to subsequent analysis and were not excluded. Mental age matched control data produced higher numbers of outliers with a single mild case observed in each of the

KERA, PoFA and DaFEx tasks, and four mild cases in the case of the KERA-Static. In all instances extreme cases could be accounted for by variation in chronological age, and were included in the final data set.

Finally, data from chronologically age matched controls produced three mild outliers in relation to the KERA, two on the NimStim and colour control task and one on the age control task. Outliers all fell at the lower end of the spectrum and were exclusively observed on scales where the interquartile rage was comparatively small. The magnitude of scores were deemed to be 'reasonable' deviations from the norm (e.g., 2.5 points below the IQR on a scale with a maximum of 22), whereby outlier status was likely exaggerated, due to the reduced variability often observed in a smaller sample sizes. Again, the identified outliers were considered 'real' results and therefore very valuable in drawing accurate conclusions. Further, as the premise of the current thesis was to reduce experimental bias apparent in the ID assessment literature, the trimming of lower bound but arguably reasonable scores in comparison groups may unduly exaggerate ID related deficits, further compounding existing experimental and ethical issues.

The detection of bivariate outliers were also made with reference to the IQR, where the notion of halfspace depth (Tukey, 1974) was applied as a generalisation of ranks to multivariate data. To visually identify outliers, bagplots (Rousseeuw, Ruts, & Tukey, 1999), the bivariate version of a boxplot, were generated using Free Statistics Software (Version 1.1.23-r7) (Wessa, 2012). A total of two bivariate outliers were detected in data derived from the index group and pertained to comparisons performed between the KERA, and the PoFA and DaFEx. Both cases were excluded from subsequent analysis.

Missing data. The proportion for missing values for the index, mental age matched and chronologically age matched control groups were, 6.6%, 1.6% and 0% respectively. Data were missing completely at random as indicated by Little's (1988) MCAR test (index group: χ^2 (167) = 22.042, p = 1.00; mental age matched control group: χ^2 (196) = 3.241, p = 1.00). When only a very small portion of data are missing (e.g., < 5% overall) and data are missing completely at random, imputation using the expectation maximization algorithm provides unbiased parameter estimates and improves statistical power (Scheffer, 2002). Missing data were imputed using Missing Values Analysis within SPSS Version 23.

Statistical assumptions of substantive analyses applied across the general linear model. Inferential data analyses were largely based on the use of parametric tests. Accordingly, statistical assumptions of normality, homogeneity of variance and linearity were considered, and are expanded upon further in the context of the specific statistical tests presented. Table 7.1 provides a reference for univariate normality data for all variables considered in the analysis. Distribution symmetry and pointedness were calculated based on z-scores for skewness and kurtosis statistics, whereby an absolute value greater than 1.96, 2.58 and 3.29 indicated statistical significance at $p < .05$, $p < .01$, and $p < .001$ respectively (Ghasemi & Zahediasl, 2012).

Table 7.1

Skewness and Kurtosis Indices for Individual Experimental Variable Distributions

Variable	Adults with ID				Mental age matched controls				Chronologically age matched controls			
	Skewness		Kurtosis		Skewness		Kurtosis		Skewness		Kurtosis	
	Statistic	z-score	Statistic	z-score	Statistic	z-score	Statistic	z-score	Statistic	z-score	Statistic	z-score
Emotion recognition task total and difference scores												
KERA Total Score	-0.34	-0.72	-0.77	-0.84	-0.79	-1.67	1.15	1.25	-1.25**	-2.64	0.26	0.28
KERA Group Total Item Score	0.28	0.60	0.10	0.11	-0.42	-0.89	-0.47	-0.50	-1.40**	-2.97	1.02	1.11
KERA-Static Total Score	-0.36	-0.77	-0.61	-0.66	-0.58	-1.22	0.95	1.03	-0.25	-0.52	-0.96	-1.05
KERA-Static Group Total Item Score	0.25	0.54	-1.35	-1.41	-0.44	-0.92	-1.03	-1.08	-0.93	-1.97	0.38	0.41
Difference score (KERA - KERA-Static)	-0.52	-1.10	0.26	0.29	-0.45	-0.96	-0.88	-0.96	0.17	0.35	-0.54	-0.58
PoFA Total Score	-0.09	-0.18	-1.20	-1.30	-1.33**	-2.82	2.52**	2.74	-0.79	-1.68	0.44	0.48
NimStim Total Score	-0.27	-0.57	-1.00	-1.09	-0.76	-1.61	-0.46	-0.50	-1.74***	-3.69	3.48***	3.79
DaFEx Total Score	-0.25	-0.52	-0.23	-0.25	-0.81	-1.72	2.45**	2.67	-0.27	-0.57	-0.21	-0.22
Emotion recognition task composite scores												
Total	-0.53	-1.13	-1.43	-1.56	-0.99*	-2.09	1.65	1.80				
Anger	-0.10	-0.21	0.02	0.02	-0.24	-0.50	-0.31	-0.34	-0.44	-0.94	-1.49	-1.62
Disgust	0.32	0.67	-1.15	-1.26	-1.30**	-2.76	1.37	1.49	-1.87***	-3.96	4.01***	4.37
Fear	-0.01	-0.01	-1.46	-1.59	0.03	0.06	0.48	0.53	-1.67***	-3.53	3.55***	3.87
Happiness	-1.72***	-3.64	2.46**	2.68	-3.34***	-7.08	12.34***	13.44	-3.07***	-6.50	9.46***	10.30
Sadness	-0.20	-0.43	-1.17	-1.27	-1.08*	-2.28	1.61	1.75	-2.57***	-5.45	8.88***	9.67
Surprise	-1.20*	-2.54	-0.01	-0.01	-1.71	-3.62	3.19***	3.47	-1.41**	-2.99	1.92*	2.09
Control tasks												
Colour discrimination	-1.16*	-2.45	0.78	0.85	-0.98	-2.07*	-1.15	-1.25	-1.41**	-2.98	1.64	1.78
Age Discrimination	-0.31	-0.66	-0.14	-0.15	-0.44	-0.93	-0.04	-0.04	-0.33	-0.70	0.44	0.47

Note. Standard error of the mean for skewness and kurtosis statistics were 0.47 and 0.92 respectively.
*$p<0.05$ **$p<0.01$ ***$p<0.001$

Table 7.1. Continued

Variable	Adults with ID				Mental age matched controls				Chronologically age matched controls			
	Skewness		Kurtosis		Skewness		Kurtosis		Skewness		Kurtosis	
	Statistic	z-score	Statistic	z-score	Statistic	z-score	Statistic	z-score	Statistic	z-score	Statistic	z-score
Social functioning measures												
DiBAS-R												
Social interaction/communication	0.10	0.21	-0.51	-0.56								
Stereotypy/rigidity/sensory	0.48	1.02	-0.28	-0.31								
DiBAS-R Total	-0.25	-0.53	-0.72	-0.79								
Autism Spectrum Quotient - Child					-0.18	-0.38	-0.23	-0.25				
Autism Spectrum Quotient - Adult									0.59	1.24	-0.53	-0.57
SPSS												
Prosocial behaviour	-0.56	-1.20	-0.03	-0.03								
Communication	-0.45	-0.96	0.04	0.04								
Appropriate assertion	0.07	0.15	0.02	0.02								
Maladaptive behaviour	0.05	0.10	-0.81	-0.88								
Sociopathic behaviour	0.07	0.15	-1.21	-1.31								
Inappropriate assertion	0.31	0.67	-0.93	-1.01								
VABS-2												
Socialisation Index	0.43	0.92	-0.10	-0.11	-0.44	-0.93	-0.16	-0.17				
Interpersonal	0.62	1.31	0.23	0.25	-0.05	-0.10	-0.99	-1.08				
Play	-0.44	-0.93	1.90*	2.07	0.24	0.51	0.40	0.43				
Coping	-0.93*	-1.98	0.79	0.86	0.67	1.42	0.01	0.01				
Participant characteristics												
Chronological age	-0.19	-0.41	-0.72	-0.79	-0.34	-0.72	-0.05	-0.05				

Note. Standard error of the mean for skewness and kurtosis statistics were 0.47 and 0.92 respectively.
*p<0.05 **p<0.01 ***p<0.001

Part 1. Evaluation of the psychometric properties of the KERA

Evaluation of reliability and reduction of items. Inter-rater reliability for the original 24 original items of the KERA were assessed based on mean total percent correct for the target emotion and Gwet's chance-corrected agreement coefficient. Reliability estimates based on responses from chronologically age matched controls are presented in Table 7.2.

Table 7.2

KERA item level inter-rater reliability estimates

Subscale/Item	Percent correct (*SE*)	AC_1 (*SE*)
Anger		
Item 1	100 (0.0)	1 (.00)
Item 2	88 (12.1)	0.76 (.13)
Item 3 †	60 (9.9)	0.37 (.12)
Item 4	76 (12.05)	0.55 (.14)
Disgust		
Item 5	96 (8.3)	0.92 (.08)
Item 6	92 (11.0)	0.84 (.12)
Item 7	88 (12.1)	0.76 (.13)
Item 8	96 (8.3)	0.92 (.08)
Fear		
Item 9	80 (12.1)	0.63 (.14)
Item 10	100 (0.0)	1 (.00)
Item 11 †	64 (8.7)	0.4 (.11)
Item 12	84 (11.3)	0.7 (.13)
Happiness		
Item 13	100 (0.0)	1 (.00)
Item 14	100 (0.0)	1 (.00)
Item 15	100 (0.0)	1 (.00)
Item 16	96 (0.0)	0.92 (.08)
Sadness		
Item 17	100 (0.0)	1 (.00)
Item 18	100 (0.0)	1 (.00)
Item 19	100 (12.1)	1 (.13)
Item 20	100 (0.0)	1 (.00)
Surprise		
Item 21	100 (0.0)	1 (.00)
Item 22	100 (0.0)	1 (.00)
Item 23	100 (0.0)	1 (.00)
Item 24	92 (12.0)	0.84 (.12)

Note. *SE* = Standard error of the mean.

† Item 3 and 11 did not meet minimum inclusion criteria and were excluded from subsequent analysis. Reliability data for Items 3 and 11 have been presented for completeness and were not included in the calculation of subscale reliability scores.

Items 3 and 11 did not meet the predetermined inclusion criteria (total percent correct 70%< and AC_1 40 <) and were excluded from the measure and subsequent analysis. Total percent correct and AC_1 for the remaining items, ranged between 76%-100% and .55-1.00 respectively. The numerical item titles of the remaining 22 items of the KERA have been reorganised so that item numbers remained consecutive.

Table 7.3 presents the KERA subscale and total scale reliability estimates with corresponding qualitative benchmarks based on Altman's scale (1991). Subscale scores fell within the moderate to very good range, with percent correct and AC_1 scores ranging from 88% (SE = 3.4) – 100% (SE = 0.0) and 0.86 (SE = 0.04) – 1.0 (SE = 0.0) respectively. Reliability estimates for the full KERA scale fell within the very good range (percent correct = 95 (SE = 1.4), AC_1 = .89 (SE = 0.03).

Table 7.3

KERA full scale and subscale reliability estimates and corresponding qualitative benchmarks

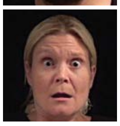
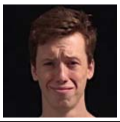
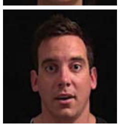
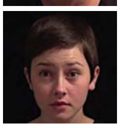
Subscale	Percent correct (*SE*)	AC_1 (*SE*)	Minimum strength of agreement achieved [a]
Anger	89 (4.3)	.78(.13)	Moderate
Disgust	93 (2.8)	.86(.04)	Very Good
Fear	88 (3.4)	.78(.12)	Moderate
Happiness	99 (1.1)	.97(.02)	Very Good
Sadness	100 (0.0)	1.0 (.00)	Very Good
Surprise	98 (1.5)	.96(.04)	Good
Total Scale	95 (1.4)	.89(.03)	Very Good

[a.] The benchmark range for agreement coefficients was calculated using Gwet's (2014) alternative benchmarking methods of cumulative membership probabilities.

Table 7.4 presents the final 22 item measure, including the demographic characteristics of the actors, item intensity ratings and an exemplar in the form a still frame extracted from the original item. Note that stills included in Table 7.4 were taken at various time points in the original clip and in many cases do not represent the maximum intensity stimuli extracted for the KERA-Static.

Table 7.4

KERA Actor Demographic Characteristics and Item Intensity Ratings

Subscale	Item exemplar	Age (Years)	Ethnicity	Item intensity
Anger				
Item1		52	Caucasian	3
Item 2		37	Caucasian	2.5
Item 3		69	Caucasian	2.7
Disgust				
Item 4		19	Caucasian	2.5
Item 5		27	Caucasian	4
Item 6		22	African (Zimba-bwean)	2.3
Item 7		49	Caucasian	2.5
Fear				
Item 8		31	Caucasian	2.5
Item 9		37	Caucasian	2.3
Item 10		44	Caucasian	1.5
Happiness				
Item 11		18	Polynesian (Māori)	4
Item 12		16	Caucasian	5
Item 13		21	Polynesian (Tongan)	3
Item 14		69	Caucasian	3.5
Sadness				
Item 15		44	Caucasian	3
Item 16		40	Caucasian	2.7
Item 17		50	Indian	3.3
Item 18		25	Caucasian	3
Surprise				
Item 19		26	Chinese	3
Item 20		27	Caucasian	2.8
Item 21		49	Caucasian	3
Item 22		16	Caucasian	2.6

Analysis of item difficulty and discrimination. The remaining 22 items of the KERA were analysed in terms of difficulty (the proportion of participants who answered an item correctly) and discriminability (how well the item serves to discriminate between participants with superior and less well developed emotion recognition abilities). Item difficulty (*p*) and discrimination (r_{pbis}) indices are included in Table 7.5.

The index group observed highly desirable overall levels of item difficulty ($p = 0.57$, $SD = 0.27$), almost exactly at the predetermined level of 0.58 required to maximize score reliability. In contrast, mental age matched and chronologically age matched controls observed higher values, 0.76 ($SD = 0.18$) and 0.95 ($SD = 0.08$) respectively, indicating reduced discriminatory power of the KERA in the two control groups. It was found that only three items in the index and mental matched groups observed extreme difficulty coefficient of 1.0, in contrast with twelve items observed by chronologically age matched controls. Extreme values observed by the index group and mental age matched controls were limited to items on the Happiness subscale, while chronologically age matched controls observed extreme performance scores across all subscales with the exception of Disgust. There were no examples of extreme difficulty coefficients of 0.0 across all three experimental groups.

Based on Ebel and Frisbie's (1991) criteria, the discriminating power of KERA items, as measured by the point-biserial correlation, fell within the very good range ($M = 0.48$, $SD = 0.23$). Overall, all four items contributing to the Happiness subscale failed to differentiate between participants of different abilities ($r_{pbis} \leq 0.20$), two items contributing to the Anger and Sadness subscales fell in the good range, and the remaining sixteen items in the very good range. Despite the low level of difficulty and discriminatory power of the Happiness subscale, the four contributing items have been preserved in the final version of the KERA, a case for which is presented in the Discussion.

Table 7.5

KERA Descriptive Statistics Including Item Discrimination and Difficulty Indices Stratified by Experimental Group

Subscale	Index Group				Mental age matched controls				Chronologically age matched controls				r_{pbis}
	M	*SD*	R	*p*	*M*	*SD*	R	*p*	*M*	*SD*	R	*p*	
Anger (max = 3)	1.21	0.98	3		1.83	0.92	3		2.67	0.64	2		
Item 1				0.54				0.71				1.00	0.66
Item 2				0.38				0.63				0.88	0.51
Item 3				0.29				0.50				0.75	0.37
Disgust (max = 4)	1.25	1.19	3		2.92	1.25	4		3.71	0.55	2		
Item 4				0.29				0.88				0.96	0.71
Item 5				0.42				0.79				0.92	0.64
Item 6				0.21				0.67				0.88	0.64
Item 7				0.33				0.58				0.96	0.61
Fear (max = 3)	0.88	1.12	3		1.8	0.93	3		2.63	0.49	1		
Item 8				0.25				0.42				0.79	0.58
Item 9				0.21				0.79				1.00	0.65
Item 10				0.42				0.58				0.83	0.54
Happiness (max = 4)	3.83	0.38	1		3.96	0.20	1		3.96	0.20	1		
Item 11				0.92				0.96				1.00	0.13
Item 12				1.00				1.00				1.00	0.0
Item 13				1.00				1.00				1.00	0.0
Item 14				1.00				1.00				1.00	0.0
Sadness (max = 4)	2.63	1.14	4		2.63	1.28	4		4	0.00	0		
Item 15				0.79				0.79				1.00	0.48
Item 16				0.75				0.63				1.00	0.39
Item 17				0.42				0.58				1.00	0.54
Item 18				0.67				0.63				1.00	0.50
Surprise (max = 4)	2.58	1.50	4		3.54	0.78	3		3.92	0.28	1		
Item 19				0.50				0.88				1.00	0.71
Item 20				0.67				0.88				1.00	0.62
Item 21				0.71				0.96				1.00	0.63
Item 22				0.71				0.83				0.92	0.47
KERA Total (max=22)	12.38	3.67	12	0.57 (*SD*=0.27)	17.15	2.94	13	0.76 (*SD*=.16)	20.83	1.45	4	0.95 (*SD*=0.08)	0.48 (*SD*=0.23)

Note. *M* = Mean, *SD* = Standard Deviation, R = Range, *p* = Item difficulty and r_{pbis} = Point Biserial Correlation

Analysis of group score distributions. Descriptive data for the overall spread of group score distributions are presented pictorially in Figure 7.1 and 7.2 and numerically in Table 7.6. When considering the KERA in its entirety, the index group observed the largest interquartile range and lowest mean and median performance scores, followed by mental age matched controls, and then chronologically matched controls. The latter observing a particularly restricted range of performances due to ceiling effects.

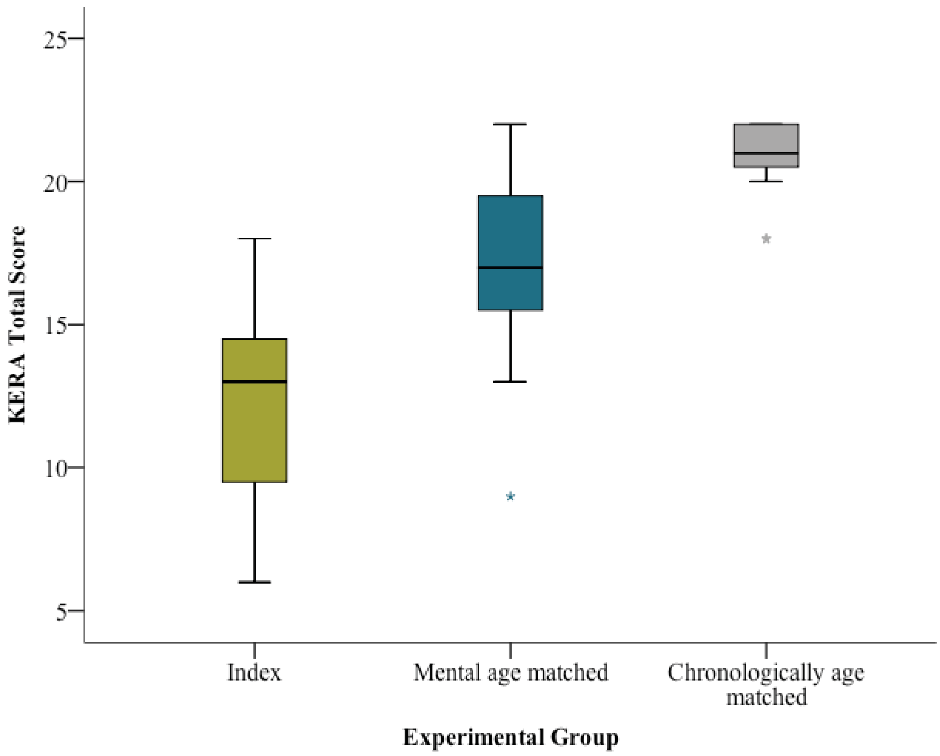


Figure 7.1. Tukey-style box plot displaying the distribution of KERA full scale scores.
Note. The KERA observes a maximum full scale score of 22.

A similar pattern of results held for performance scores at the subscale level, with the exception of Happiness scale, where all three experimental groups observed ceiling effects. However, much overlap existed between the index group and the mental age control group, and similarly between mental age matched and chronologically matched controls, the statistical significance of which will be explored in the Part 2 of the analysis.

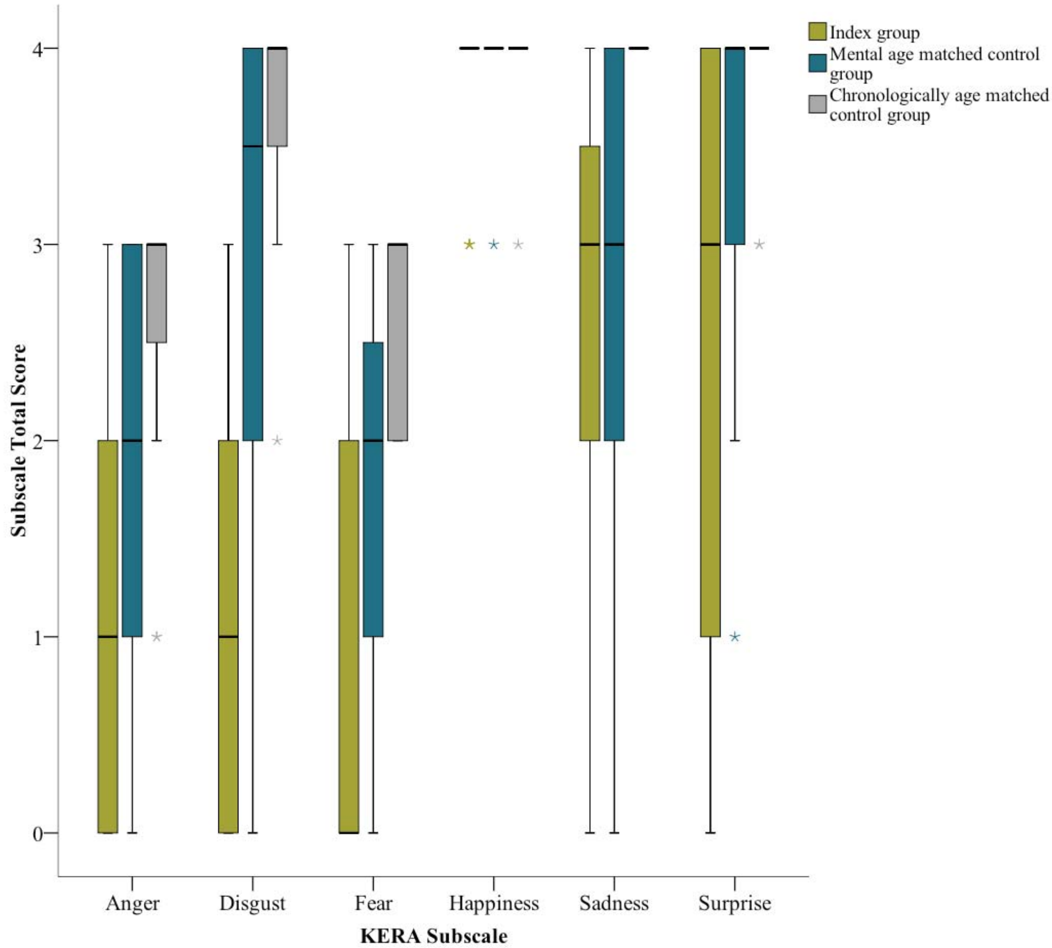


Figure 7.2. Tukey-style box plot displaying the distribution of KERA subscale scores.

Note. The KERA observes a maximum subscale scores of three for Anger and Fear, and four for Disgust, Happiness, Sadness and Surprise.

Normality data assessing distribution symmetry and pointedness are presented in Table 7.6. Skewness and kurtosis values for index group data fell within acceptable bounds of normality across the full scale and all subscales, with the exception of Happiness which was negatively skewed. Mental age and chronologically age matched controls observed increasing levels of skew and kurtosis likely to be function of ceiling effects in the two groups. The dominant pattern was that of negative skewness and positive kurtosis.

Table 7.6

KERA Skewness and Kurtosis Values Stratified by Group.

Subscale	Index Group		Mental age matched controls		Chronologically age matched controls	
	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
Anger	0.17	-1.04	-0.01	-1.15	-1.79***	2.20*
Disgust	0.32	-1.44	-0.71	-0.62	-1.80***	2.68**
Fear	0.88	-0.69	-0.25	-0.72	-0.55	-1.86**
Happiness	-1.91***	1.79	-4.90***	24.00***	-4.90***	24.00***
Sadness	-0.53	-0.34	-0.71	-0.45	-	-
Surprise	-0.55	-1.22	-1.96***	4.02***	-3.22***	9.12***
KERA Total	-0.34	-0.77	-0.79	1.15	-1.25**	0.26

Note. Standard error of the mean for skewness and kurtosis statistics were 0.47 and 0.92 respectively. Z-scores of >1.96, >2.58 and >3.29 indicated statistical significance at p < .05, p < .01, and p < .001 respectively.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Validity assessment. To assess convergent validity, Pearson product-moment correlations were calculated to examine the relationship between the KERA full scale score and the PoFA, NimStim and DaFEx. Preliminary analyses based on visual inspection of scatterplots, showed a linear relationship for index group scores, between the KERA and the three convergent validity measures. A linear relationship was also observed in control group data with the exception of DaFEx scores where there was no visually discernible pattern. Data were largely normally distributed across the three groups, and deviations from normality were limited to three examples of negative skew and positive kurtosis across the two control groups (see Table 7.1).

Opinion is varied regarding the requirement of normally distributed data when applying significance testing to Pearson's correlations. Generally, the literature indicates the Pearson correlation to be robust to normality violations (Field, 2005) with the exception of extremely non-normal distributions (e.g., Bradley L-Shaped distribution e.g., Skew ≤ 3, kurtosis =17 [Blair & Lawson,1982] which have been associated with inflated Type 1 error [Bishara & Hittner, 2012]). Based on these parameters, correlational analysis applying Pearson's correlation was deemed appropriate for the observed distribution.

Table 7.7 presents the correlation coefficients depicting the relationship between the KERA and the PoFA, NimStim and DaFEx. Pearson product-moment correlations showed high levels of convergence between the KERA and the comparison measures. Associations were largely positive and statistically significant, with the exception of control group comparisons between the KERA and the DaFEx. Based on Cohens' (1988) criteria, correlations observed in the index group were all large in magnitude, and statistically significant relationships observed in the control groups were in the moderate range.

Table 7.7

Pearson Correlations Between the KERA and the PoFA, NimStim and DaFEx.

Stimuli	Index group	Mental age matched controls	Chronologically age matched controls
PoFA	.51*	.45*	.45*
NimStim	.67**	.41*	.41*
DaFEx	.61**	.04	.04

* $p < .05$. ** $p < .01$. (two-tailed)

Simple linear regression was applied to assess the predictive validity of the KERA. Specifically, how performance on the KERA predicted autism spectrum disorder trait levels. Preliminary analyses based on visual inspection of scatterplots, showed no evidence of linearity between the KERA and autism screening measures for chronologically age matched controls (AQ-Adult) and the group were excluded from the analysis. Absence of linearity in the chronologically age matched control group is likely due to ceiling effects on the KERA attenuating potential correlations. A linear relationship between the KERA and autism screening measures were observed in both the index group (DiBAS-R) and mental age matched control group (AQ-Child). Homoscedasticity and the normal distribution of residuals were confirmed based on visual inspection of standardised residuals versus standardised predicted values plots and normal probability plots.

The prediction equation for the index group was DiBAS-R total score = 26.57 + (-.762 × KERA total score). Average KERA score statistically significantly predicted autism disorder trait levels ($F(1, 22) = 5.60$, $p < .05$) accounting for 20% of the variation in DiBAS-R scores with an adjusted R^2 of 16.7%, a medium effect size according to Cohen (1988). An increase performance score on the KERA of one-point lead to a 0.8 point drop in the DiBAS-R total score.

The prediction equation for the mental age matched control group was AQ-Child total score = 97.293 + (-3.088 x KERA total score). Average KERA score statistically significantly predicted autism disorder trait levels ($F(1, 22) = 5.60$, $p < .05$) accounting for 32.9% of the variation in AQ-Child scores with adjusted R^2 of 29.8%, a large effect size according to Cohen (1988). An increase performance score on the KERA of one-point lead to a 3.1 point drop in the AQ-Child total score.

Part 2. Investigating the nature of the emotion recognition abilities of people with ID

Part 2 statistical procedures are organised according to four research objectives: (1) Re-evaluate the applicability of the Emotion Specificity Hypothesis for individuals with Intellectual Disability; (2) Determine the effect of dynamic cues on facial emotion recognition performance in people with ID relative to typically developing individuals; (3) Investigate the effects of emotion intensity on emotion recognition performance in adults with ID relative to mental and chronologically age matched controls; and (4) Explore the link between social adaptive functioning and emotion recognition abilities in adults with ID relative to mental age matched controls.

Objective 1: Re-evaluate the applicability of the Emotion Specificity Hypothesis for individuals with intellectual disability. A series of one-way repeated measures ANOVAs were conducted to determine whether there were statistically significant differences in emotion recognition performance across the three experimental groups. Emotion recognition performance was assessed based on total scores on the KERA in addition to the PoFA, NimStim and DaFEx.

Preliminary analysis assessed data for normality and sphericity. The assumption of normality is often cited as necessary for statistical significance testing using a one-way repeated measures ANOVA. However, so long the assumption of normality is not severely violated and the design is balanced (i.e., there are an equal number of observations in each cell), the actual Type I error rates approximate nominal rates for both t-tests and analysis-of-variance type tests. The exception here is severely platykurtic data, whereby the effect of kurtosis on power increases proportionately with the presence of variables affected by kurtosis (Glass & Stanley, 1970; Stevens, 2002).

Normality data for the current study are included in Table 7.1. A small amount of negative skew was observed on the colour control task across all three experiment groups. Mental age matched control data observed negative skew on the PoFA and leptokurtosis across the PoFA and DaFEx. Chronologically age matched control data produced negative skew on the KERA and NimStim, and leptokurtosis on the NimStim. The application of repeated measures ANOVA was considered appropriate given that the majority of data points approximated a normal distribution, and in cases where data deviated, distributions were all skewed in a similar manner. Further, there were no cases of platykurtic data.

In contrast with assumptions of normality, one-way repeated measures ANOVA demonstrate extreme sensitivity to violations of sphericity, and in practice violations are difficult to avoid (Weinfurt, 2000). Furthermore, common tests of sphericity such as the Mauchly's test, often fail to detect departures from sphericity in small samples (Keselman, Rogan, Mendoza, & Breen, 1980). Accordingly, results from all repeated measures ANOVA calculations were interpreted using a Greenhouse-Geisser correction (1959) to adjust degrees of freedom for within subjects variables and error effect.

Table 7.8 summarises the descriptive statistics and repeated measure ANOVA results, comparing task performance across the three experimental groups. Mean performance scores are presented pictorially in Figure 7.3.

Table 7.8

One-Way Repeated Measures ANOVA Results Comparing Group Effects for Experimental Tasks

Stimuli	Index group		Mental age matched controls		Chronologically age matched controls		F	df	Effect size (ω^2)
	M	*SD*	*M*	*SD*	*M*	*SD*			
KERA	12.4	3.7	17.2	2.9	20.9	1.5	68.90***	1.9, 42.5	0.65
PoFA	10.3	3.4	13.7	2.5	15.8	1.8	34.50***	1.8, 41.2	0.48
NimStim	11.5	3.1	16.0	2.0	17.0	1.3	44.38***	1.6, 36.5	0.55
DaFEx	10.4	2.5	13.1	2.1	16.0	1.3	52.34***	1.9, 43.2	0.59
Age discrimination task	7.7	1.6	9.1	1.3	9.5	1.5	11.30***	42.2, 85.8	0.22
Colour discrimination task	11.2	0.9	11.7	0.5	11.3	0.9	2.71	1.8, 42.3	0.05

Note. M = Mean, *SD* = Standard Deviation
* $p < .05$. ** $p < .01$. *** $p < .001$. With Greenhouse-Geisser correction applied.

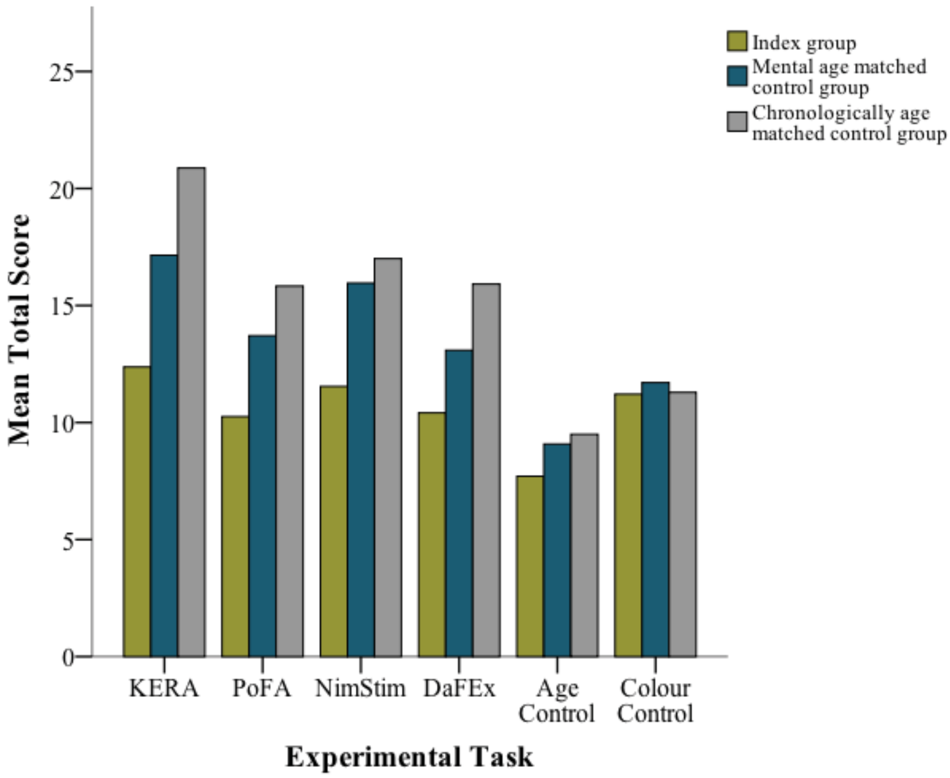


Figure 7.3. Emotion recognition task total performance scores organised by experimental group. *Note.* The KERA observes a maximum full scale scores of 22, the PoFA, NimStim and DaFEx a total score of 18, and the control tasks a total score of 12.

The one-way repeated measures ANOVA revealed a statistically significant main effect of group membership across all tasks with the exception of the colour discrimination control task. Between 48-65% of variation observed in performance scores on the emotion recognition tasks, and 22% on the age discrimination task, may be explained by group membership. To isolate specific significant group differences, post hoc analysis included multiple paired-samples t-tests with Bonferroni adjustment for multiple comparisons. Priori analysis showed difference scores to be normally distributed. Table 7.9 presents the pairwise comparisons of task performance across experimental group.

Table 7.9

Pairwise Comparisons of Task Performance Across Experimental Group

Stimuli	Pair	Mean Difference	95% Confidence Interval for Difference	Effect size (Cohen's d)
KERA	MA-I	4.80***	2.83 – 6.76	1.44
	CA-I	8.50***	6.46 – 10.54	3.32
	CA-MA	3.71***	2.12 – 5.30	1.70
PoFA	MA-I	3.46***	1.52 – 5.40	1.18
	CA-I	5.58***	3.74 – 7.43	2.15
	CA-MA	2.13*	0.70 – 3.55	0.98
NimStim	MA-I	4.42***	2.72 – 6.12	1.78
	CA-I	5.46***	3.61 – 7.30	2.54
	CA-MA	1.04	0.08 – 2.17	0.62
DaFEx	MA-I	2.67***	1.14 – 4.20	1.19
	CA-I	5.50***	4.10 – 6.90	3.00
	CA-MA	2.83***	1.61 – 4.05	1.75
Age discrimination task	MA-I	1.38**	0.39 – 2.36	0.97
	CA-I	1.80**	0.62 – 2.97	1.18
	CA-MA	0.42	0.46 – 1.30	0.29

Note. I = Index Group, MA = Mental age matched control group, and CA = Chronologically age matched control group.
* $p < .05$. ** $p < .01$. *** $p < .001$. Adjusted for multiple comparisons using the Bonferroni correction.

The prevailing pattern was that of superior performance by the chronologically age matched control group, followed by the mental age matched control group and the index group. Pairwise comparisons supporting this pattern were significant in all instances, with the exception of group differences observed on the NimStim and age discrimination task for the two control groups. The mean effect size observed on emotion recognition tasks between the index group and the mental and chronologically age matched control groups was d = 1.31 and d = 2.44 respectively. All significant difference scores produced large effect sizes.

The described pattern of scores, is largely in keeping with the ESH, with the exception of the significant main effect observed between the index group and control groups on the age discrimination task (control task). Nevertheless, it should be noted that the effect sizes observed on the age discrimination task were the lowest observed, with the exception of comparable results between the index group and mental age matched controls on the PoFA. Full support for the Emotion Specificity Hypothesis would have required a significant performance discrepancy between the index and control groups on tasks of emotion recognition and no differences in group performance across both control measures.

To determine whether observed emotion recognition deficits were emotion specific or global in nature, an additional series of analysis-of-variance type tests were completed. Subscale scores were collapsed across the KERA, PoFA, NimStim and DaFEx and served as the dependent variables. Priori analysis showed that control group data demonstrated extreme deviations from the normal distribution and nonparametric tests were selected. The Friedman test was applied to determine whether any statistically significant differences existed between the three groups in regard to their performance across discrete emotion categories.

Mean group performance scores for individual emotion categories are presented in Figure 7.4. Again, the prevailing pattern across all emotion categories was that of superior performance by the chronologically age matched control group, followed by the mental age matched control group and weakest performance by the index group. This pattern was preserved across each subscale of the KERA, PoFA, NimStim and DaFEx**.** Individual task subscale scores are presented in Figure J.1. (Appendix J).

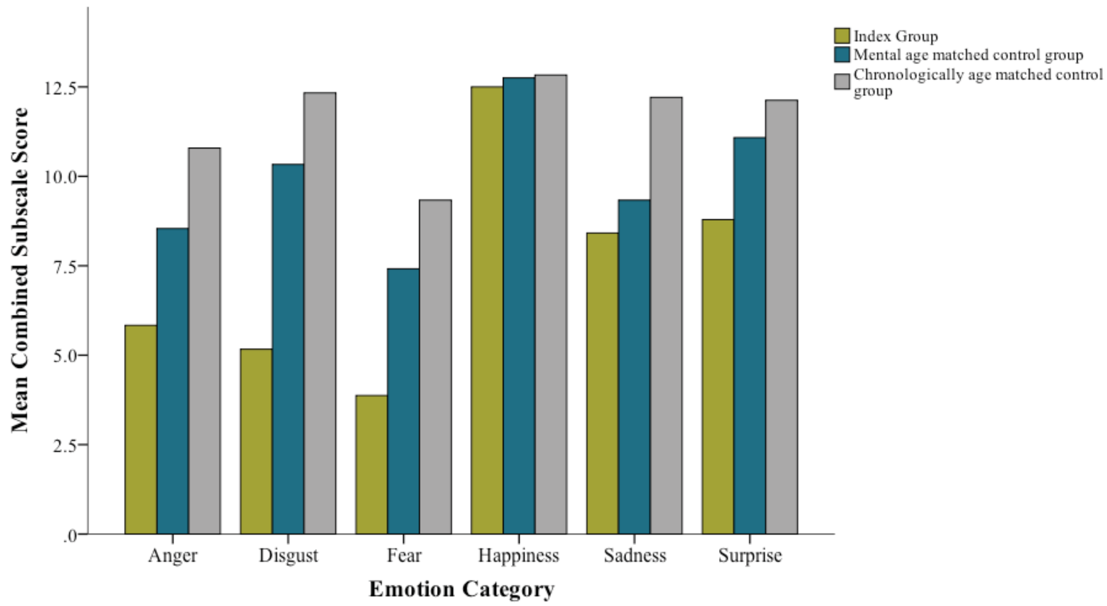


Figure 7.4. Group emotion recognition task performance organised by emotion category.

Table 7.10 summarises the descriptive statistics and results of the Friedman test, comparing task performance across the three experimental groups. The Friedman test revealed a statistically significant main effect across all emotion categories with the exception of Happiness. Kendall's W effect sizes ranged from weak to strong indicating that the magnitude of observed group differences were dependent on emotion category.

Table 7.10

Friedman Test Results Comparing Group Effects for Emotion Specific Task Performance

Stimuli	Index group		Mental age matched control group		Chronologically age matched control group		χ^2 (2)	Effect size (Kendall's W)
	Mdn	*IQR*	*Mdn*	*IQR*	*Mdn*	*IQR*		
Anger	5.5	3.8	9.0	3.00	11.0	2.8	29.30***	0.61
Disgust	4.5	6.8	11.5	3.8	13.0	1.0	34.14***	0.71
Fear	4.0	5.0	7.0	2.8	10.0	3.0	32.28***	0.67
Happiness	13.0	1.0	13.0	0.0	13.0	0.0	3.211	0.07
Sadness	9.0	4.8	9.5	1.8	12.0	1.0	27.22***	0.57
Surprise	11.0	5.5	12.0	1.75	12.0	1.0	12.72**	0.27

Note. *Mdn*=Median, *IQR*=Interquartile Range, and χ^2= chi squared distribution.
* $p < .05$. ** $p < .01$. *** $p < .001$. Adjusted for multiple comparisons using the Bonferroni correction.

To isolate specific group differences, post hoc analysis included a series of Wilcoxon signed-rank tests with Bonferroni correction. Priori analysis showed difference scores to be normally distributed. Table 7.11 presents pairwise comparisons of emotion specific task performance across experimental group.

In increasing order of magnitude, the index group showed significant performance deficits when compared with the mental age matched control group on scales measuring anger, disgust and fear. Similarly, comparisons between the index group and the chronologically age matched control group, demonstrated index group deficits on scales measuring surprise, sadness, fear, anger and disgust. When comparing mental age matched and chronologically age matched controls, the later observed significantly higher scores on measures of anger and sadness. All significant comparisons resulted in large effect sizes indicating a strong association between group membership and emotion level performance outcomes.

Table 7.11

Pairwise Comparisons of Emotion Category Task Performance Across Experimental Group

Stimuli	Pair	Mean Difference (SE = 0.29)	Effect size (r) [a]
Anger	MA-I	0.79***	0.67
	CA-I	1.52***	0.86
	CA-MA	0.73***	0.72
Disgust	MA-I	1.02***	0.81
	CA-I	1.60***	0.88
	CA-MA	0.58	0.56
Fear	MA-I	0.96*	0.81
	CA-I	1.60***	0.85
	CA-MA	0.65	0.54
Happiness	MA-I	-	-
	CA-I	-	-
	CA-MA	-	-
Sadness	MA-I	0.21	0.29
	CA-I	1.35***	0.84
	CA-MA	1.15***	0.80
Surprise	MA-I	0.33	0.49
	CA-I	0.92**	0.64
	CA-MA	0.58	0.55

Note. I = Index Group, MA = Mental age matched control group, and CA = Chronologically age matched control group; *SE* = Standard error.
[a.] $r = Z/\sqrt{N}$ (Rosenthal, 1994)
* $p < .05$. ** $p < .01$. *** $p < .001$. Adjusted for multiple comparisons using the Bonferroni correction.

Objective 2: Determine the effect of dynamic cues on facial emotion recognition performance in people with ID, relative to typically developing individuals. To determine the effect of dynamic cues on facial emotion recognition as a function of group membership, three paired samples t-tests with Bonferroni correction were used to compare group mean performance scores on the KERA and the KERA-Static. Preliminary analysis based on skewness and kurtosis indices (see Table 7.1) showed the distribution of within group differences between the KERA and KERA-Static to be normally distributed

Table 7.12 presents the results of the within group pairwise comparisons of task performance on the KERA and KERA-Static. A statistically significant mean difference was observed among the control groups, whereby performance on the KERA was superior to performance on the KERA-static. Effect sizes fell within the large range suggesting high practical significance. In contrast, no significant difference were observed across the two test forms for the index group.

Table 7.12

Within Group Pairwise Comparisons of Task Performance on the KERA and KERA-Static

Group	KERA		KERA Static		Paired difference		95% Confidence Interval for Difference	T (23)	Effect size (Cohen's d)
	M	*SD*	*M*	*SD*	*M*	*SD*			
Index group	12.38	3.67	12.46	2.32	-0.08	8.86	-1.29 – 1.12	-0.14	0.01
Mental age matched	17.15	2.94	15.80	2.57	1.38	2.70	0.24 – 2.52	2.50*	0.51
Chronologically age matched	20.88	1.45	17.62	2.30	3.25	2.09	2.37 – 4.13	7.62***	1.56

Note. *M* = Mean, *SD* = Standard Deviation.
* $p < .05$. ** $p < .01$. *** $p < .001$. (Two -tailed). Adjusted for multiple comparisons using the Bonferroni correction.

Objective 3. Investigate the effects of emotion intensity on emotion recognition performance in adults with ID relative to typically developing individuals. Pearson correlation coefficients observed between the KERA and KERA-Static item total scores and item intensity, were compared across each experimental group. To establish individual item intensity, each of the FACS AU intensity codes were assigned a numeric value: trace, 1; slight evidence, 2; marked or pronounced, 3; severe or extreme, 4; and maximum evidence, 5. A global item intensity score was calculated for each individual item by dividing the sum of intensity codes by the number of contributing AUs. Overall item intensity ratings ranged from 1.5-5 with a mean of 3.0 (*SD* = 0.71).

Preliminary analysis showed the KERA and KERA-Static item total scores to be normally distributed across the three experimental groups (see Table 7.1). The visual inspection of scatterplots confirmed a linear relationship across all comparisons except in instances where there was no visually discernible pattern. Table 7.13 presents correlation coefficients summarising the relationship between emotion intensity and emotion recognition accuracy. A positive relationship was observed across all comparisons. The index group observed large significant effects across both the KERA and KERA-Static, while mental age matched controls observed moderate effect sizes where only scores on the KERA-Static showed statistical significance. The chronologically age matched control group observed small to medium effect sizes however the relationship observed was not statistically significant.

Table 7.13

Pearson Correlation Coefficients Between Emotion Intensity and Group Emotion Recognition Performance

Stimuli	Index group	Mental age matched control group	Chronologically age matched control group
KERA	.59**	.40	.12
KERA- Static	.51*	.43*	.33

* $p < .05$. ** $p < .01$. *** $p < .001$. (two-tailed)

Objective 4. Explore the link between social adaptive functioning and emotion recognition abilities in adults with ID relative to mental age matched controls. Total scores were combined across the KERA, PoFA NimStim and DaFEx, to form an emotion recognition composite score which was then correlated with measures of social functioning. Both the index group and mental age matched control group composite scores were compared against the VABS-2 Socialisation Index. Index group composite scores were also compared against the

SPSS and DiBAS-R Social Communication and Interaction subscale. To examine potential gender effects separate analyses were run for males and females. In addition to the full correlation between the emotion recognition composite and measures of social functioning, partial correlations were performed to control for the effects of chronological age.

Preliminary analysis based on visual inspection of scatterplots confirmed a linear relationship across all comparisons except in instances where there was no visually discernible pattern. Data approximated a normal distribution with very mild negative skew and positive kurtosis across index group scores on the VABS-2 Coping and Play scales respectively, and negative skew on the emotion recognition composite for mental age matched controls. Participant chronological age was normally distributed within each experimental group (see Table 7.1). Observed deviations from normality were few in number and relatively mild, indicating that the statistical significance of Pearson's correlation coefficients could be safely interpreted.

Descriptive data are summarised in Table 7.14. Emotion recognition performance and social adaptive functioning as measured by the VABS-2, were notably higher in the control group. The various measures of social adaptive functioning demonstrated comparable results across males and females in both experimental groups.

Table 7.14

Group Mean Performance Scores on the Emotion Recognition Composite and Measures of Social Adaptive Functioning

Stimuli	Index group			Mental age matched control group		
	Total sample	Male	Female	Total sample	Male	Female
Emotion recognition composite	44.6 (10.5)	42.0 (11.2)	46.13 (10.15)	59.9 (6.9)	59.6 (6.2)	60.1 (7.6)
VABS-2 Socialisation Index	53.5 (19.9)	53.8 (24.8)	53.4 (17.4)	119.3 (11.8)	126.6 (10.5)	114.9 (10.5)
Interpersonal relationships	6.8 (3.9)	6.6 (4.5)	6.9 (3.6)	18.2 (3.1)	20.9 (1.8)	16.6 (2.5)
Play and Leisure Time	8.7 (3.7)	9.7 (3.7)	8.0 (3.7)	18.1 (2.2)	18.1 (1.8)	18.1 (2.4)
Coping skills	9.4 (3.7)	9.6 (2.9)	9.2 (4.1)	18.3 (2.5)	17.8 (1.2)	18.6 (3.0)
SPSS						
SPSS Maladaptive total	29.51 (16.1)	27.4 (16.6)	30.85 (16.3)			
Sociopathic behaviour	13.7 (7.8)	13.4 (8.5)	13.7 (7.6)			
Inappropriate assertion	15.9 (9.7)	14.0 (9.2)	17.0 (10.2)			
SPSS Prosocial total	65.1 (13.2)	58.9 (14.7)	69.1 (10.9)			
Communication	43.4 (8.4)	39.1 (8.8)	46.2 (7.2)			
Appropriate Social Skills	21.7 (5.9)	19.8 (6.8)	22.9 (5.3)			
DiBAS- R Communication and Interaction	11.7 (4.4)	12.51 (4.2)	10.89 (4.0)			

Note. Standard deviation indicated in parentheses.

Table 7.15 presents Pearson's correlation coefficients summarising the relationship between emotion recognition and social adaptive functioning outcomes. Few significant relationships were observed between emotion recognition and social functioning for both experimental groups. A number of moderate correlations were observed in the index group total sample data; however, these were not preserved when age effects were accounted for. Statistically significant gender specific relationships that remained after controlling for age were limited to females in the index group, who observed a strong positive relationship between the emotion recognition composite score and the VABS-2 Socialisation Index and Play and Leisure Time subscale. Across the total sample, the direction of observed relationships was as anticipated, in that indices of adaptive social behaviour were positively correlated with emotion recognition performance, and maladaptive behaviour indices were negatively correlated. The exception being small arguably negligible correlations (.05>). Note that for the DiBAS-R Communication and Interaction subscale, higher scores are associated with poorer social performance.

Table 7.15

Pearson Correlation Coefficients Between the Emotion Recognition Composite Score and Measures of Social Adaptive Functioning

Stimuli	Index group			Mental age matched controls		
	Total sample	Male	Female	Total sample	Male	Female
VABS-2 Socialisation Index	.47* (.40)	.37 (.02)	.58* (.69***)	.16 (.26)	.22 (.25)	.21 (.43)
Interpersonal relationships	.27 (.02)	.18 (.36)	.33 (.18)	.04 (.13)	-.13 (-.08)	.17 (.39)
Play and Leisure Time	.24 (.15)	-.01 (-.28)	.50 (.55*)	-.04 (.09)	-.01 (.12)	-.05 (.07)
Coping skills	.36 (.78)	.14 (-.05)	.50 (.44)	.26 (.39)	-.03 (-.27)	.37 (.59)
SPSS Maladaptive subscale total	-.12 (.03)	.06 (.05)	-.27 (.00)			
Sociopathic behaviour	-.11 (.05)	.07 (.06)	-.25 (.08)			
Inappropriate assertion	-.10 (.02)	.04 (.03)	-.24 (-.05)			
SPSS Prosocial subscale total	.42* (.36)	.55 (.40)	.24 (.23)			
Communication	.41* (.38)	.57 (.46)	.23 (.19)			
Appropriate Social Skills	.34 (.26)	.46 (.24)	.18 (.21)			
DiBAS-R Communication and Interaction Subscale	-.23 (-.09)	.32 (.60)	-.61* (-.52)			

Note. Partial correlations with the effect of age removed are shown in parentheses.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$. (two-tailed)

It was hypothesised that the low incidence of significant correlations and modest effect sizes observed, may be attributed to the sample size (particularly in regard to male participants) coupled with the potentially low ecological validity of the emotion recognition composite score. Within the composite score only half of the contained tasks included dynamic stimuli and the PoFA items did not include colour cues. Consequently, there is possibility that the composite was not the most accurate approximation of real-world emotional expressions. To test this hypothesis, specifically whether a stronger association might be observed as a function of relying exclusively on dynamic stimuli, the analyses was run a second time substituting the KERA and KERA-Static total scores in place of the emotion recognition composite. As previously mentioned, priori analysis demonstrated that KERA and KERA-Static scores to be normally distributed across the two experimental groups (see Table 7.1).

Table 7.16. summarises mean KERA and KERA-Static total scores organised by group membership and gender. Emotion recognition performance as measured by the KERA and KERA-Static, were notably higher in the control group and females performed marginally better than males across all comparisons.

Table 7.16

Group Mean Performance Scores on the KERA and KERA-Static

Stimuli	Index group			Mental age matched control group		
	Total sample	Male	Female	Total sample	Male	Female
KERA	12.40 (3.7)	11.67 (4.2)	12.8 (3.4)	17.2 (2.9)	16.9 (2.8)	17.3 (3.1)
KERA-Static	12.46 (2.32)	12.44 (2.2)	12.5 (2.4)	15.8 (2.6)	14.3 (2.7)	16.67 (2.1)

Note. Standard Deviation indicated in parentheses.

Table 7.17 and 7.18 present correlation coefficients summarising the relationship between the KERA and KERA-Static and social adaptive functioning outcomes. Consistent with hypotheses, a larger number of significant relationships were observed based on the KERA scores when compared with both the emotion recognition composite and the KERA-Static. Results derived from the dynamic form are used here to make inferences regarding the link between social functioning and emotion recognition ability, for the reason that the dynamic form may be considered the more valid measure.

Overall, the anticipated positive relationship between social functioning and emotion recognition was most apparent in the index group. Moderate to large correlations were observed in the total sample across the VABS-2 Socialisation index and associated subscales. Observed correlations were preserved after controlling for age with the exception of the Interpersonal Relationships subscale. This same pattern of results was demonstrated in the female subgroup, though all effect sizes fell within the large range. In regard to the remaining social adaptive functioning scales, a single strong negative correlation was observed for the female subgroup on the DiBAS Communication and Interaction subscale.

Table 7.17

Pearson Correlation Coefficients Between the KERA Total Score and Measures of Social Adaptive Functioning

Stimuli	Index group			Mental age matched control group		
	Total sample	Male	Female	Total Sample	Male	Female
VABS-2 Socialisation Index	.57** (.51*)	.57 (.27)	.58* (.77**)	.35 (.38)	.17 (.17)	.58* (.70**)
Interpersonal relationships	.42* (.19)	.24 (-.44)	.58* (.49)	.05 (.07)	-.18 (-.19)	.25 (.32)
Play and Leisure Time	.48* (.47*)	.42 (.22)	.62** (.77**)	.09 (.13)	-.12 (-.16)	.17 (.22)
Coping skills	.52** (.47*)	.41 (.24)	.62** (.64*)	.38 (.42*)	.03 (.03)	.47 (.55*)
SPSS Maladaptive subscale total	-.33 (-.20)	-.13 (-.17)	-.51 (-.31)			
Sociopathic behaviour	-.32 (-.18)	-.12 (-.15)	-.50 (-.25)			
Inappropriate assertion	-.28 (-.18)	-.12 (-.16)	-.44 (-.30)			
SPSS Prosocial subscale total	.36 (.30)	.31 (-.03)	.37 (.40)			
Communication	.36 (.32)	.28 (.03)	.40 (.40)			
Appropriate Social Skills	.30 (.24)	.32(-.11)	.23 (.29)			
DiBAS Communication and Interaction Subscale	-.23 (-.07)	.38 (-.07)	-.72** (-.66)			

* $p < .05$. ** $p < .01$. *** $p < .001$. (two-tailed)
Note. Partial correlations with the effect of age removed are shown in parentheses.

Table 7.18

Pearson Correlation Coefficients Between the KERA-Static Total Score and Measures of Social Adaptive Functioning

Stimuli	Index group			Mental age matched controls		
	Total sample	Male	Female	Total sample	Male	Female
VABS-2 Socialisation Index	.28 (.23)	.18 (-.00)	.36 (.44)	-.10 (-.07)	.35 (.36)	.01 (.14)
Interpersonal relationships	.11 (-.07)	-.13 (-.58)	.26(.15)	-.40 (-.38)	-.04 (-.03)	-.19 (-.09)
Play and Leisure Time	.20 (.18)	.26 (.18)	.18(.20)	-.07 (-.03)	.24 (.34)	-.26 (-.19)
Coping skills	.44* (.45*)	.62 (.60)	.38 (.39)	.22 (0.28)	-.21 (-.20)	.31 (.45)
SPSS Maladaptive subscale total	-.35 (-.28)	-.71* (-.73*)	-.14 (.02)			
Sociopathic behaviour	-.33 (-.26)	-.55 (-.57)	-.20 (-.02)			
Inappropriate assertion	-.31 (-.26)	-.77* (-.79*)	-.08 (.04)			
SPSS Prosocial subscale total	.21 (.152)	-.19 (-.42)	.55 (.55*)			
Communication	.22 (.18)	-.28 (-.45)	.62* (.62*)			
Appropriate Social Skills	.14 (.08)	-.77 (-.28)	.30 (.31)			
DiBAS Communication and Interaction Subscale	-.33 (-.26)	.21 (.33)	-.71** (-.67)			

* $p < .05$. ** $p < .01$. *** $p < .001$. (two-tailed)
Note. Partial correlations with the effect of age removed are shown in parentheses.

Mental age matched data observed a significant positive partial correlation for the total sample and female subgroup on the VABS-2 Coping skills subscale, in addition to a strong total and partial correlation for the female subgroup on the VABS-2 Socialisation Index. Similar to results based on the emotion recognition composite score, almost all correlations observed across the two experimental groups were in the predicted direction. There was one exception within the mental age matched male data, a small non-significant effect on the VABS-2 Interpersonal relationships and Play and Leisure Time subscale.

CHAPTER 8

DISCUSSION

Overview

The current study sought to clarify the nature of the emotion recognition abilities of people with ID and relevant outcomes in social functioning. In the interests of developing a fair assessment of emotion recognition appropriate for use with this group, the Kinetic Emotion Recognition Assessment (KERA) was developed and piloted. The KERA was then applied in the assessment of the ESH and emotion recognition moderating factors, including movement and emotion intensity. To explore the link between emotion recognition ability and social competence, emotion recognition data collected in the course of measure development and the assessment of the ESH was also interpreted in the context of three measures of social functioning. This chapter opens with a summary of the psychometric properties achieved by the KERA, followed by findings relating to the ESH, emotion recognition moderating factors, and the relationship between emotion recognition and social functioning. Potential limitations and suggestions for future research are presented, followed by an executive summary to capture the substance and scope of what has been attempted by this research, including contributions to research and clinical implications.

Summary of study aims and findings

The development and psychometric assessment of the Kinetic Emotion Recognition Assessment (KERA). The first aim of the study was to develop and pilot the Kinetic Emotion Recognition Assessment. The KERA is a brief 22-item measure including colour real-time videotaped expressions of basic emotions and a forced choice response format including six response options. The KERA also offers a wide range of FACS validated emotion intensities.

The development and psychometric evaluation of the KERA was established in two stages. The first included the construction of an item pool and selection of a core set, and the second, the pilot testing of the measure and item refinement. Content validity procedures applied during the initial development stage are first described, followed by a summary of item level and full scale psychometric properties as assessed in the subsequent pilot study.

Content validity. The content validity of the KERA is strongly supported by the stringent reliability and validity criteria applied during the item development and selection phase of the study. The full corpus was first reduced to 50 clips based on emotion recognition accuracy and naturalness rankings by nine independent raters. A core set of 24 items was then extracted based on independent EMFACS review.

Consensus scoring by independent raters was deemed a valid method to reduce the initial pool of items, based on the underlying assumption of discrete emotion theories regarding the evolutionary and social foundations of emotion and emotional expression (Ekman & Cordaro, 2011). In this regard, high levels of rater agreement (≥ 70% agreement) not only serve as preliminary evidence of inter-rater reliability but are also interpreted as evidence toward content validity. Naturalness ratings were included in the interests of face validity. The subsequent application of EMFACS served as a ground truth or additional content validity check, to ensure that all potential items put forward for the KERA were based on well-established action unit code combinations associated with specific emotions. Finally, assessment of individual FACS codes (the bases for EMFACS) also served as a replicable basis from which to derive global emotion intensity ratings.

Inter-rater reliability, item difficulty and discrimination, and criterion validity. The second stage of development involved pilot testing the measure and included: the assessment of item reliability and subsequent item refinement, followed by evaluation of subscale and full scale inter-rater reliability; assessment of item difficulty and discrimination and analysis of score distributions; and finally, an evaluation of the convergent and predictive validity of the measure.

Inter-rater reliability. Reliability data were not derived from individuals with ID, as emotion recognition capacities could not be assumed to be intact. Instead, data were based on chronologically age matched adult controls. Item and subscale reliability was assessed using identical parameters set for inter-rater agreement, applied in the preliminary selection of clips from the original corpus. Only two items did not meet predefined criteria, and were excluded from the measure and subsequent analysis. Overall a high level of inter-rater reliability was achieved in the final measure. Reliability estimates for the full scale KERA fell within the very good range, and subscale scores within the moderate to very good range.

These findings corroborate the results of the inter-rater agreement assessment conducted in the initial phases of item development. In addition, considering the previously established content validity of the core item set, the diversity of inter-reliability scores may be assumed to represent differences in item difficulty, due to item intensity, rather than variation in the level of validity across the different items. Results to verify this statement are included in the 'Emotion Intensity' subsection of this discussion.

Item difficulty, discrimination and total score distributions. Aligned with the central aims of the study, the index group observed highly desirable overall levels of item difficulty, almost exactly at the predetermined level required (p = 0.58) to maximize score reliability. In contrast, mental age matched and chronologically age matched controls observed higher values, indicating reduced discriminatory power of the KERA in the two control groups. It was found that only three items in the index and mental matched groups observed extreme difficulty coefficient of 1.0, in contrast with 12 items observed by chronologically age matched controls. Extreme values observed by the index group and mental age matched controls were limited to items on the Happiness subscale, while chronologically age matched controls observed extreme performance scores across all subscales with the exception of Disgust. There were no examples of extreme difficulty coefficients of 0.0 across any of the experimental groups, indicating that there no items on the KERA served to directly restrict the variability of test scores as a result of being too challenging.

The extreme high scores observed on items on the Happiness subscale are in keeping with findings from previous studies for both ID (Rojahn, Rabold, & Schneider, 1995; Wishart, Cebula, Willis, & Pitcairn, 2007) and neurotypical populations (Calvo et al., 2014). It is recognised that due to ceiling effects, these items are unlikely to contribute to measuring individual differences for any of the three groups studied. However, task performance for individuals with ID is more adversely impacted by experiences of failure, when compared with typically developing individuals (Weisz, 1999). Accordingly, all happiness items were preserved in the final measure in the hope that these items could also serve to bolster test taker confidence in addition to maintaining the full range of basic emotions. Easily achievable items may also serve as an indicator for a lack of test understanding, or low effort or persistence. Low effort or persistence is particularly relevant here as it is linked to lower levels of 'expectancy of success' (Weisz, 1999), which is observed at higher rates in individuals with ID (Roy, Retzer, & Sikabofori, 2015; Zigler, Bennett-Gates, Hodapp, & Henrich, 2002).

Following reliability analysis, the discriminating power of test items was assessed. As we would expect based on the previous assessment of item difficulty indices, only the four items contributing to the Happiness subscale failed to differentiate between participants of different abilities ($r_{pbis} \leq 0.20$). These findings are a direct result of the ease at which people recognise the emotion happiness, and the associated items were preserved in the final measure for reasons previously discussed. Overall, highly desirable levels of item discriminatory power were demonstrated across the KERA. Based on Ebel and Frisbie (1991) criteria, sixteen of the 18 remaining items fell within the very good range, and two items contributing to the Anger and Sadness subscales fell in the good range.

Building on individual item analysis, an examination of group score distributions was completed to assess the appropriateness of the KERA for use with adults with ID when compared with the two control groups. Consistent with the central aims of the study, individuals with ID observed a wide spread of normally distributed full scale scores with minimal floor and ceiling effects. Individuals with ID also demonstrated the largest interquartile range when compared with the two control groups and the mean score for individuals with ID fell approximately at 50%. A similar pattern of results held for performance scores at the subscale level, with the exception of Happiness scale, where all three experimental groups observed ceiling effects. Skewness and kurtosis indices for the group ID group data fell within acceptable bounds of normality across the full scale and all subscales, again with the exception of Happiness. In contrast, mental age and chronologically age matched controls observed increasing levels of skew and kurtosis due to ceiling effects in the two groups.

Aligned with the central goals of this thesis, the findings from item difficulty and discrimination analysis in combination with analysis of group score distributions, suggest that the KERA is pitched at the appropriate level to capture the full range of emotion recognition capacities expected of individuals with nonspecific mild ID. Further, all items (except those comprising the Happiness subscale) discriminate between test takers with high and low scores. As one might expect, the KERA is however not suited for use with typically developing adults due to ceiling effects restricting the range of possible scores. Utility with typically developing children falls somewhere between individuals with ID and typically developing adults, and the KERA is most likely to be suited to younger children with lesser developed emotion recognition abilities than the sample applied here (mean = 9 years and 6 months).

Convergent and predictive validity. The results offer strong evidence in favour of the convergent validity of the KERA. A high degree of convergence was observed between the KERA and the comparison measures (PoFA, NimStim and DaFEx) across all three experimental groups. Pearson product-moment correlations were positive and statistically significant, with the exception of control group comparisons between the KERA and the DaFEx. Significant relationships observed for participants with ID were all large in magnitude, while statistically significant relationships observed in the control groups were in the moderate range. The reduced levels of convergence observed in the control groups is likely explained by ceiling effects restricting variance at the upper end of the performance spectrum.

The absence of a positive relationship between the DaFEx and the KERA for the two control groups remains curious, given that the DaFEx was the only comparison measure that, like the KERA, incorporated dynamic cues. There are two plausible explanations for these findings. Either the DaFEx and or the KERA suffer limitations in construct validity, the latter being less likely in light of the favourable psychometric properties already demonstrated. During the measure development phase of this study, the DaFEx was considered the best available video-based measures to assess convergent validity. Nevertheless, like all measures, the DaFEx bears limitations. First, while inter-rater reliability has been assessed for the DaFEx no-utterance condition applied in this study, estimates (excluding the neutral condition) averaged 75%, considerably lower than the 95% achieved by the KERA. Further, item level data for the DaFEx is not reported in evaluation studies, and the specific reliability estimates for the selected DaFEx stimuli remain unspecified. Finally, evaluation studies to date have relied exclusively on consensus scoring and have not included assessment of criterion validity, or measures of content validity such as the assessment of facial action units. Without such verification, it is difficult to draw firm conclusions as to why the KERA and DaFEx did not produce the predicted relationship. Future enquiry might apply the ADFES-BIV, a FACS verified dynamic stimuli set published during the data collection phase of the current study.

Finally, the current study offers preliminary evidence for the predictive validity of the KERA. Consistent with literature linking greater levels of emotion recognition ability with lower levels of autism traits (Poljac et al., 2012; Uljarevic & Hamilton, 2013), the KERA accounted for substantial variance in traits of autism for both individuals with ID (16.7%) and typically developing child controls (29.8%). Differences observed between the participants with ID and child controls could not be interpreted, due to the application of separate autism screening measures. It may simply be that the longer AQ-Child captured more overall variance when compared with the DiBAS-R. Finally, a threshold for the ideal proportion of explained variance for both groups could not be established, due to a lack of available research employing correlational designs upon which to inform hypotheses. Nevertheless, the medium to large effect sizes observed are in keeping with reasonable expectations.

Investigating the nature of emotion recognition abilities in people with ID. The objectives of the second part of the study were to: re-evaluate the applicability of the emotion specificity hypothesis for individuals with Intellectual Disability; identify potential moderating effects of dynamic cues and emotion intensity on emotion recognition performance; and investigate the link between social adaptive functioning and emotion recognition abilities in adults with ID and typically developing children.

Emotion specificity hypothesis. Full support for the ESH demands two criteria are met. First, there must be evidence of a significant performance discrepancy between participants with ID and control groups on tasks of emotion recognition. Second, participants with ID and control participants must exhibit comparable performance across control measures void of emotional content but with similar task related demands. The latter is crucial in determining whether potential emotion recognition difficulties observed, are secondary to the cognitive limitations inherent in ID, or better explained by the ESH as domain specific facial emotion processing deficits.

Consistent with the ESH, when compared with mental and chronologically age matched controls, participants with ID demonstrated relative impairment in emotion recognition as measured by the KERA, PoFA, NimStim and DaFEx. Dependent on the emotion recognition task applied, group membership explained between 48% and 65% of the variation observed in performance scores. The large magnitude of the observed effect sizes are in keeping with existing literature (for a recent review see Scotland, Cossar, & McKenzie, 2015).

Findings also suggest that when compared with individuals with similarly developed cognitive abilities (i.e., the mental age matched control group), relative impairment observed by people with ID is limited to a subset of emotion categories (anger, disgust and fear). In contrast, differences between people with ID and same age neurotypical peers seem to be more global, with the exception of happiness. Assessment of potential differences in recognition accuracy for happiness was limited by ceiling effects across all three experimental groups. Nonetheless, the study confirms that people with ID possess at least a basic aptitude for recognising happiness. The majority of the previously discussed ESH studies (see Table 3.1) report results in a holistic manner, and do not detail specific emotion level differences. Consequently, interpretation of emotion specific findings in reference with other studies is difficult. The current findings are however aligned with Rojahn, Rabold, et al. (1995) key study, whereby performance for happiness items were comparable and differences in sadness between individuals with ID and child controls were arguably negligible.

The current study offers strong evidence of a relative impairment in emotion recognition for individuals with ID when compared with both mental and chronologically age matched peers. However, whether observed difficulties are domain specific remains less clear due to inconsistent performance across the two control tasks. Participants with ID performed comparably to controls on the colour discrimination task; however all three groups demonstrated ceiling effects, limiting opportunity to detect potential group differences. Nevertheless, the results offer important insight regarding the specificity of difficulties demonstrated by participants with ID. First, the high level of performance observed on the colour discrimination task affords confidence that participants were well equipped to manage basic task demands. Specifically, attending to dynamic colour stimuli of brief duration and selecting an appropriate response from six available options.

In contrast with performance on the colour discrimination task, participants with ID were outperformed by both control groups on the age discrimination task. Full support for the ESH would dictate that individuals with ID demonstrate an exclusive deficit on the emotion tasks, and comparable performance with the control groups across the two controls tasks. Accordingly, at surface level relative deficits on the age discrimination task might suggest that instead of a domain specific emotion processing deficits, individuals with ID possess a more general deficit in the processing of human faces. However, observation taken during testing suggests that impaired performance may be, at least in part, due to unanticipated difficulties managing more general task related demands. Specifically, the numerical age ranges accompanying the qualitative descriptors 'young adult', 'middle aged adult' and 'old adult'. While all participants were able to grasp the age groupings at the outset of the task, many did not appear to retain this information for the entire task duration and required prompting throughout. It is also important to note that the effect sizes based on mean difference scores for the age discrimination task were the lowest observed of all the experimental tasks that produced significant group differences.

In summary, the current study provides strong evidence that individuals with ID experience relative impairment in emotion recognition abilities when compared with typically developing controls. There is some evidence that impairment is limited to a select number of emotions when individuals are matched for developmental level and are more global in nature when comparisons are based on chronological age. Findings gleaned from the control tasks afford confidence that impairment is not due to cognitive limitations associated with entertaining dynamic colour stimuli of brief duration and selecting an appropriate response from six available options. However, akin to limitations reported in previous studies, there was some evidence that the age discrimination control task was not equally matched in terms of task related demands. Consequently, the relative poor performance of individuals with ID on this task presents two possible interpretations. First, we may consider these findings evidence against the ESH, specifically that impairment is limited to the general processing of faces and not specific to emotions. Alternatively, in light of the relatively small effect sizes observed for group differences on the age discrimination task when compared with the emotion recognition task, it is equally plausible that an emotion specific deficit does exist and poor performance on the age discrimination task was simply an artefact of overly complex response options.

The effect of dynamic cues and intensity on facial emotion recognition performance. A heavy focus of the current study was also to explore how subtle methodological differences in the operationalisation of facial emotion may affect emotion recognition performance. Specifically, whether the presence of ID moderated the reported beneficial effects of dynamic cues and higher levels of emotion intensity. It was hoped that the investigation would offer insight regarding how to best interpret existing ESH literature, for which typical stimuli include static images of unverified intensity.

Dynamic cues. The effect of dynamic cues on facial emotion recognition performance, were examined by comparing group total performance scores on the KERA and the KERA-Static. Application of static items derived from original dynamic video clips provided a high degree of experimental control across a wide range of variables including subject age, gender and ethnicity. Consistent with previous literature, dynamic cues were found to bolster performance for neurotypical individuals (Alves, 2013). Effect sizes fell within the large range suggesting high practical significance.

Curiously, participants with ID performed comparably on static and dynamic versions of the KERA, suggesting that unlike neurotypical controls this group do not benefit from the addition of dynamic cues. These findings are at odds with results reported in a similar study by Harwood, Hall and Shinkfield (1999), who found that dynamic cues enhanced overall performance for both groups. An obvious point of difference between the two studies was the stimuli duration. In contrast with the KERA where clips ranged from 2-10 seconds and were terminated at the natural apex of the emotion, all stimuli in the Harwood et al., (1999) study were of fixed duration (10 seconds). A description of the stimuli reported that clips were terminated at the completion of the facial expression, and static images extracted from the apex. Many emotions, such as surprise manifest rather rapidly (e.g., 330ms; Yoshikawa & Sato, 2007), therefore it is more likely that the natural apex occurred much earlier in the clip than the selected endpoint. It may be then that the static images in the Harwood et al., (1999) study did not include the extra information conveyed by the maximal facial muscle displacement that occurs at the natural apex of emotion. In this case, increased levels of accuracy observed by individuals with ID when rating dynamic stimuli may not have been a result of motion cues, but simply because the dynamic clips included the natural apex of the emotion.

Emotion intensity. In keeping with previous literature, typically developing participants observed a positive relationship between emotion intensity and emotion recognition performance (Hess, Blairy, & Kleck, 1997; Hoffmann, Kessler, Eppel, Rukavina, & Traue, 2010; Montagne, Kessels, De Haan, & Perrett, 2007). Mental age matched participants observed moderate effect sizes; however, only scores on the static version of the KERA demonstrated statistical significance. These results may be explained in the context of Bould and Morris' (2008) findings, where the benefit of motion cues for neurotypical individuals was reduced with higher intensity expressions. Accordingly, we may expect a diminished relationship between dynamic stimuli and emotion recognition performance, due to dynamic stimuli bolstering performance exclusively for lower intensity items. Consistent with this hypothesis, chronologically age matched controls observed a similar pattern of results, where the correlation was reduced for dynamic items. Lack of statistical significance observed by this group for both the static and dynamic versions of the KERA, is likely due to overall high performance restricting variance.

Participants with intellectual disability observed a positive relationship between emotion intensity and emotion, with comparable effect sizes across both the KERA and KERA-Static. These findings are consistent with literature based on the general population, and in light of findings that suggest individuals with ID do not benefit from dynamic cues it was anticipated that we would not expect to see the relationship offset on by an interaction effects between dynamic cues and emotion intensity.

These results are, however, inconsistent with Gray, Fraser, and Leudar (1983), who reported that individuals with mild ID do not benefit from increasing emotion intensity. The two studies differ in regard to their underlying conceptual framework, and the discrepant results may be explained in terms of the distinct methods by which emotion intensity is operationalised. Gray, and colleagues (1983) were informed by Schlosberg and Harold's (1954) theory of emotion, whereby emotion is organised according to dimensions of intensity, pleasantness versus unpleasantness, and attention versus rejection. Discrete emotions are then mapped onto the three-dimensional space. For example, happiness is placed low along the intensity dimension and high on the pleasantness dimension. Within this paradigm intensity refers to overall level of activation of the organism and is considered fixed within each emotion category. While overall physiological activation is likely to map onto the intensity of facial movements, this model does not capture differences of intensity within discrete emotion categories. A case can then be made that by capturing relative intensity within individual emotion categories, the current study offers a more nuanced approach to assessing the effects of facial intensity affording confidence in the current findings.

Emotion recognition and social adaptive functioning. The final goal of the current study was to explore the link between facial emotion recognition and social performance. It was hoped that emotion recognition ability may be identified as a key factor in social adaptive functioning and therefore an appropriate target for intervention.

Total scores were combined across the KERA, PoFA NimStim and DaFEx to form an emotion recognition composite score which was then correlated against various measures of social functioning. After controlling for chronological age, few significant relationships were observed between emotion recognition and social functioning for both individuals with ID and mental age matched controls. It was anticipated that these findings reflected the poor ecological validity of the emotion recognition composite score, on account that only half of the contributing tasks included dynamic stimuli. One could not expect the composite score to correlate highly with real-world behavioural outcomes, if the score did not reflect an accurate approximation of real-world emotional expressions. To test this hypothesis, correlations were also computed for the KERA and KERA-Static.

As expected, a considerably larger number of significant relationships were observed, based on the KERA scores when compared with both the emotion recognition composite and the KERA-Static. This finding alone offers strong evidence that static stimuli do not offer an ecologically valid assessment tool, and offers a likely explanation for why just two (Rojahn, Esbensen, & Hoch, 2006; Rojahn & Warren, 1997) of the six existing studies (García-Villamisar, Rojahn, Zaja, & Jodra, 2010; Simon, Rosen, Grossman, & Pratowski, 1995; Williams et al., 2005; Wishart et al., 2007), demonstrated a significant link between facial emotion recognition ability and social adaptive functioning for people with ID.

In light of these findings, only results derived from the dynamic form of the KERA are used to make inferences regarding the link between social functioning and emotion recognition ability. Overall, the study findings were consistent with the notion that facial emotion recognition abilities are linked to outcomes in social functioning. A positive relationship was observed between emotion recognition and pro-social functioning across all the applied measures. The positive relationship between social functioning and emotion recognition was most pronounced for individuals with ID. Given that the ID group displayed lesser developed emotion recognition abilities compared to controls, this may suggest a type of threshold effect where emotion processing deficits only disrupt social adaptive functioning beyond a certain level of severity. Curiously, the only statistically significant comparisons observed across both groups were for female participants. A lack of statistical significance for male participants may be a result of reduced statistical power due to the smaller number of participants in this group. However, stronger links between social functioning and emotion recognition for females have been reported within the general population, and therefore it is equally plausible that this is just further evidence that the ability to identify others' emotional states from nonverbal cues is a relatively more important socio-cognitive ability for females when compared with males (Custrini & Feldman, 1989; Leppänen & Hietanen, 2001).

Regarding specific components of social functioning linked to emotion recognition performance, female control participants only observed statistically significant results on the VABS-2 Socialisation index as a whole, while female participants with ID also observed significant effects on the Play and Leisure and Coping Skills subscales of the VABS-2. Together these scales measure how the individual plays, shares, and cooperates, recognises social cues, and demonstrates responsibility and sensitivity toward others.

The focus of the current study was on positive social behaviours; however, maladaptive behaviour was also assessed, due to the integrated format of the SPSS measure meaning that it was best presented in its entirety. A negative relationship was observed between emotion recognition and indices of inappropriate assertion and sociopathic behaviour. These results are unsurprising being that reduced insight regarding the emotional states of others may mean that individuals are likely to continue with antisocial or irksome behaviours due to difficulty interpreting non-verbal feedback.

Limitations and considerations for future research

The current study fulfilled the research aims, however there are some limitations. Specifically, difficulty developing control tasks of appropriate complexity and sample related factors. This section summarises potential limitations and how these may be improved upon in future research. The successful pursuit of the original research questions has also highlighted new avenues for future research which are also presented.

Limitations. The primary limitation of this study, was related specifically to the assessment of the ESH. Akin to previously published studies, there was difficulty with selecting a face based control task of comparable complexity to the emotion recognition tasks. Some participants with ID experienced difficulty grasping the numerical age ranges accompanying the assigned response options on the age discrimination task. Consequently, insight regarding the specificity of observed emotion recognition deficits remains limited. Chronological age, like emotion, is one of the few physical human attributes where small changes in form result in a significant change in meaning. Unfortunately, age does not lend itself to be divided into distinct categories with clear boundaries. Future research may overcome this issue by presenting the response options in a simpler format, such as a sliding scale with the anchors 'young' and 'old'.

The current study could have been further strengthened by an increase in sample size. While the sample size was in keeping with the priori power analysis and previous ESH literature, a larger participant pool would have afforded more statistical power and confidence in our findings. This is particularly pertinent in instances where the absence of an effect resulted in conclusions disparate with previous research, as was demonstrated in our investigation of movement cues. In contrast, demographic sample composition, one of the most frequently cited threats to external validity, is not considered a major limitation in this study. This study is founded upon Basic Emotion Theory which suggests that primary human emotions are innate and universal, indicating that fundamental recognition processes should be generalisable across all groups. Further, application of matched pairs designed offered additional assurance that demographic factors (cultural gender and age differences) did not bias results.

Future directions. Aside from the general replication of findings, the most pressing issue for future research is to establish a standardisation sample for the KERA, from which group norms may be established. This research may be extended to include alternative groups with documented emotion recognition difficulties such as autism spectrum disorder and Down syndrome (Wishart, Cebula, Willis, & Pitcairn, 2007). In addition, assessment of the temporal stability of the KERA would improve the utility of the measure in assessing change over time in the context of emotion recognition skills intervention.

This study offers strong evidence in favour of the ESH; however, support was not unequivocal and it remain unclear whether observed deficits are specific to the general processing of faces or emotion cues. Future research would benefit from continued enquiry regarding the specificity of the observed impairments in emotion recognition for people ID. This may be achieved through replication of the current study adjusting for the identified issues in the response format of the age discrimination control task. Such a study may be bolstered by including formal intelligence testing for all participants, so that individual cognitive capacities may be explicitly controlled for. While participants in the current study had undergone cognitive assessment, score profiles were not available for the purposes of this study.

Complete support for the ESH would naturally raise the question as to why individuals with ID demonstrate an emotion recognition deficit. The current study offers a promising starting point, with preliminary findings suggesting that unlike their typically developing peers, individuals with ID may not benefit from movement cues. Future research should serve to replicate these findings, while controlling for general cognitive processes such as processing speed. A second avenue may be to investigate factors associated with poor emotion recognition in different developmental conditions. For example, poor emotion recognition for individuals with autism is associated with a reliance on local featural processing rather than global configural processing of faces (Harms, Martin, & Wallace, 2010). Further supporting this line of enquiry is research that suggests there may be a possible overlap in genetic aetiology with non-syndromic ID and autism spectrum disorder (Ellison et al., 2013; Kaufman et al., 2010). Preliminary studies have begun to investigate these factors (e.g., Scotland, McKenzie, Cossar, Murray, & Michie, 2016); however, to the author's knowledge such studies have relied on the processing of static objects rather than faces.

A final avenue of enquiry for future research, would be to assess for causality between emotion recognition performance and social functioning. The goals of this study were somewhat exploratory, in that limited literature existed to inform predictions regarding the direction of tested relationships or whether a relationship did exist. Accordingly, correlational analyses were deemed appropriate for these purposes. Nevertheless, correlation does not imply causality and these assumptions must be tested explicitly. To this end, a longitudinal study involving targeted intervention of emotion recognition skills is required.

Executive summary

Intellectual disability is associated with varying degrees of impairment in social adaptive functioning, which serve to negatively impact quality of life through social isolation, increased incidence of mental health difficulties and stigmatization (Matson & Hammer, 1996). Deficits in facial emotion recognition are often cited as a key component skill responsible for the social difficulties experienced by people with ID (Rojahn, Rabold, et al., 1995). This position has been formally conceptualised by the emotion specificity hypothesis (ESH; Rojahn, Rabold, & Schneider, 1995), which proposes that individuals with ID manifest a specific deficit in facial emotion recognition beyond that which can be explained by difficulties in general intellectual functioning. Informed by the ESH, several well-meaning researchers recommended that social intervention programmes target emotion recognition skills (Mcalpine, Singh, Ellis, Kendall, & Hampton, 1992; Owen, Browning, & Jones, 2001). These recommendations were deemed premature, as studies demonstrating a link between emotion recognition ability and social adaptive functioning were few, and the validity of the ESH was undermined by mixed aetiological samples and limitations in the ecological validity of existing emotion stimuli. The overarching goal of the current study was to address these limitations in the literature applying a bottom up approach. First, by clarifying the nature of emotion recognition deficits in the ID population, followed by assessment of relevant outcomes in social adaptive functioning.

To this end the KERA was developed in the interests of improving on the ecological validity of existing measures applied in ESH research. The KERA is the first measure validated for use with nonspecific ID, that includes FACS verified dynamic expressions of varying intensity. Initial psychometric evidence supported the inter-rater reliability and overarching construct validity of the measure, offering strong evidence in favour of content and convergent validity. The KERA also accounted for substantial variance in traits of autism, supporting the predictive validity of the measure. Finally, individual item analysis also suggested that the KERA is pitched at the appropriate level of difficulty to capture the full range of emotion recognition capacities expected of individuals with nonspecific mild ID. While the inspiration for the KERA developed from a desire to improve assessment of the ESH, the potential applications of the measure extend well beyond the scope of this study.

Within the current study, the KERA served as a reliable and ecologically valid basis upon which to base investigation of emotion recognition abilities in ID and relevant social outcomes. While unequivocal support for the ESH was not established, due to issues demonstrating domain specificity, the findings offer strong evidence that individuals with ID experience relative impairment in emotion recognition abilities when compared with typically developing controls. Investigation regarding the impact of emotion intensity and movement cues offers insight into observed impairment. While preliminary, the findings suggest that similar to their typically developing peers, individuals with ID benefit from higher intensity emotional displays. Curiously however, they do not benefit from the addition of movement cues. These findings would suggest that previous studies that have applied exclusively static stimuli, are likely to have underestimated relative emotion recognition impairment for individuals with ID when compared with typically developing individuals. These findings also speak to the utility of the KERA in emotion research, being that it is the only measure designed in the interests of people with nonspecific ID that also incorporates validated dynamic stimuli.

Perhaps the most significant contribution of the current study are findings that emotion recognition ability is in fact linked to social adaptive functioning. This is an important finding given that purpose of research is ultimately to improve real-world outcomes. These findings give hope that social intervention programmes focused on enhancing emotion recognition skills will ultimately lead to improved social functioning. In addition, observation of a relatively stronger association between social functioning and dynamic stimuli when compared with static images, suggests the latter do not offer an ecologically valid representation of emotion. Stimuli applied in the course of intervention should therefore include real time displays of emotion such as those offered by the KERA, in the interests of maximising the generalisability of acquired skills.

The current study represents significant progress in both clarifying the emotion recognition abilities of people with nonspecific ID, and in improving the standards of experimental enquiry applied with this often overlooked group. The findings have produced interesting insights into the nature of emotion recognition impairment and relevant outcomes in social functioning. Most importantly the study identifies potential mechanisms underlying emotion recognition difficulties and offers a valid tool with which to assess emotion recognition capacities. This is an exciting area for future research, as we move beyond simply identifying impairment toward developing empirically guided interventions.

REFERENCES

Adolphs, R. (2002). Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and Cognitive Neuroscience Reviews*, *1*(1), 21–62. doi:10.1177/1534582302001001003

Ainsworth, P., & Baker, P. C. (2004). *Understanding Mental Retardation*. Mississippi, MS: University Press of Mississippi.

Altman, D. G. (1991). *Practical Statistics for Medical Research*. Retrieved from https://books.google.co.nz/books?id=vwalRnRxWQC&printsec=frontcover&dq=Altman+D.+Practical+Statistics+for+Medical+Research.&hl=en&sa=X&redir_esc=y#v=onepage&q=Altman D. Practical Statistics for Medical Research.&f=false

Alves, N. T. (2013). Recognition of static and dynamic facial expressions: A study review. *Estudos de Psicologia (Natal)*, *18*(1), 125–130. doi:10.1590/S1413-294X2013000100020

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*(5), 403–410. doi:10.1111/j.0956-7976.2005.01548.x

American Psychiatric Association. (2013). *Statistical Manual of Mental Disorders* (5th Ed.). Washington: D.C: American Psychiatric Association: Author.

Apple Inc. (2013). Apple - Final Cut Pro X (Version 10.1.1)[software]. Retrieved from https://itunes.apple.com/us/app/final-cut-pro/id424389933?mt=12

Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The autism spectrum quotient: Children's version (AQ-Child) [Measurement Instrument]. *Journal of Autism and Developmental Disorders*, *38*(7), 1230–1240. doi:10.1007/s10803-007-0504-z

Barbato, G., Barini, E. M., Genta, G., & Levi, R. (2011). Features and performance of some outlier detection methods. *Journal of Applied Statistics*, *38*(10), 2133–2149. doi:10.1080/02664763.2010.545119

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5-17. doi:10.1023/A:1005653411471

Barr, C. L., & Kleck, R. E. (1995). Self-other perception of the intensity of facial expressions of emotion: Do we know what we show? *Journal of Personality and Social Psychology*, *68*(4), 608–618. doi:10.1037/0022-3514.68.4.608

Bate, S., & Bennetts, R. (2015). The independence of expression and identity in face-processing: Evidence from neuropsychological case studies. *Frontiers in psychology*, *6*, 1-7. doi:10.3389/fpsyg.2015.00770

Battocchi, A., Pianesi, F., & Goren-Bar, D. (2005). A first evaluation study of a Database of Kinetic Facial Expressions (DaFEx). *ICMI '05: Proceedings of the 7th International Conference on Multimodal interfaces*, 214–221. doi:10.1145/1088463.1088501

Beauchamp, M. H., & Anderson, V. (2010). SOCIAL: An integrative framework for the development of social skills. *Psychological Bulletin*, *136*(1), 39–64. doi:10.1037/a0017768

Bedell, J. R., & Lennox, S. S. (1997). *Handbook for communication and problem-solving skills training: A cognitive-behavioral approach*. New York, NY: John Wiley & Sons.

Benson, J. B., & Haith, M. M. (2009). *Diseases and disorders in infancy and early childhood*. Oxford, England: Academic Press.

Bielecki, J., & Swender, S. L. (2004). The assessment of social functioning in individuals with mental retardation: A review. *Behavior Modification*, *28*(5), 694–708. doi:10.1177/0145445503259828

Bisch, J., Kreifelts, B., Bretscher, J., Wildgruber, D., Fallgatter, A., & Ethofer, T. (2016). Emotion perception in adult attention-deficit hyperactivity disorder. *Journal of Neural Transmission*, *123*(8), 961–70. doi:10.1007/s00702-016-1513-x

Bishara, A. J., & Hittner, J. B. (2012). Testing the significance of a correlation with nonnormal data: Comparison of Pearson, Spearman, transformation, and resampling approaches.

Psychological Methods, *17*(3), 399–417. doi:10.1037/a0028087

Blair, R. C., & Lawson, S. B. (1982). Another look at the robustness of the product-moment correlation coefficient to population non-normality. *Florida Journal of Educational Research*. *24*, 11-15. Retrieved from http://psycnet.apa.org/psycinfo/1983-20165-001

Bould, E., & Morris, N. (2008). Role of Motion Signals in Recognizing Subtle Facial Expressions of Emotion. *British Journal of Psychology*, *99*(2), 167–189. doi.org/10.1348/000712607X206702

Bould, E., Morris, N., & Wink, B. (2008). Recognising subtle emotional expressions: The role of facial movements. *Cognition & Emotion*, *22*(8), 1569–1587. doi:10.1080/02699930801921156

Bray, A. (2003). *Definitions of intellectual disability: Review of the literature prepared for the National Advisory Committee on Health and Disability to inform its project on services for adults with an intellectual disability*. Retrieved from Donald Beasley Institute website: http://www.donaldbeasley.org.nz/publications/NHC_Definitions.pdf

Brosgole, L., Gioia, J. V, & Zingmond, R. (1986). Facial- and postural-affect recognition in the mentally handicapped and normal young children. *The International Journal of Neuroscience*, *30*(1–2), 127–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3744708

Buntinx, W., Cobigo, V., McLaughlin, C., Morin, D., Tassé, M. J., & Thompson, J. R. (2008). *Psychometric Properties of the Supports Intensity Scale* [White paper]. Retrieved from American Association on Intellectual and Developmental Disabilities website: https://aaidd.org/docs/default-source/sis-docs/siswppsychometric.pdf?sfvrsn=0

Burack, J. A., Hodapp, R. M., Iarocci, G., & Zigler, E. (2012). On Knowing More: Future Issues for Developmental Approaches to Understanding Persons with Intellectual Disability. In Author (Eds.), *The Oxford Handbook of Intellectual Disability and Development* (pp. 395-402). New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780195305012.013.0025

Burack, J. A., Russo, N., Gordon Green, C., Landry, O., & Iarocci, G. (2016). Developments in the developmental approach to intellectual disability. In D. Cicchetti (Ed.), *Developmental psychopathology: Maladaption and psychopathology* (3rd ed., pp. 1–67). New Jersey, NJ: John Wiley & Sons.

Butcher, L. M., Meaburn, E., Knight, J., Sham, P. C., Schalkwyk, L. C., Craig, I. W., & Plomin, R. (2005). SNPs, microarrays and pooled DNA: Identification of four loci associated with mild mental impairment in a sample of 6000 children. *Human Molecular Genetics*, *14*(10), 1315–1325. doi:10.1093/hmg/ddi142

Calvo, M. G., Gutiérrez-García, A., Fernández-Martín, A., Nummenmaa, L., Calvo, M. G., Gutiérrez-García, Á. A., … Nummenmaa, L. (2014). Recognition of facial expressions of emotion is related to their frequency in everyday life. *Journal of Nonverbal Behavior*, *38*, 549–567. doi:10.1007/s10919-014-0191-3

Carr, A. (2016). *Handbook of Child and Adolescent Clinical Psychology: A Contextual Approach* (3rd ed.). New York: Routledge.

Carrow-Woolfolk, E. (1995). *Comprehensive Assesssment of Spoken Language*. Circle Pines, MN: American Guidance Services.

Cheal, J. L., & Rutherford, M. D. (2011). Categorical perception of emotional facial expressions in preschoolers. *Journal of Experimental Child Psychology*, *110*(3), 434–443. doi:10.1016/j.jecp.2011.03.007

Chung, K., Reavis, S., Mosconi, M., Drewry, J., Matthews, T., & Tassé, M. J. (2007). Peer-mediated social skills training program for young children with high-functioning autism. *Research in Developmental Disabilities*, *28*(4), 423–436. doi:10.1016/j.ridd.2006.05.002

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (end ed.). Retrieved from https://www.amazon.com/Statistical-Power-Analysis-Behavioral-Sciences/dp/0805802835

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. doi:10.1037/0033-2909.112.1.155

Crick, N. R., & Dodge, K. A. (1994). Review and reformulation of social information-processing mechanisms in chidren's social adjustment. *Psychological Bulletin*, *115*(1), 74–101. doi:10.1037/0033-2909.115.1.74

Custrini, R. J., & Feldman, R. S. (1989). Children's social competence and nonverbal encoding and decoding of emotion. *Journal of Clinical Child Psychology*, *18*(4), 336–342. doi:10.1207/s15374424jccp1804_7

Darwin, C. (1998). The expression of the emotions in man and animals: *Introduction, afterword, and commentaries by Paul Ekman*. P. Ekman (Ed.). Oxford, England: Oxford University Press. (Original work published 1872)

Davies, B. E., Frude, N., Jenkins, R., Hill, C., & Harding, C. (2015). A study examining the relationship between alexithymia and challenging behaviour in adults with intellectual disability. *Journal of Intellectual Disability Research*, *59*(11), 1022–1032. doi:10.1111/jir.12186

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions, Psychological Science, *11*(1), 86–89. doi: 10.1111/1467-9280.00221

Dobson, E., & Rust, J. O. (1994). Memory for objects and faces by the mentally retarded and nonretarded. *Journal of Psychology*, *128*(3), 315. doi: 10.1080/00223980.1994.9712735

Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test, Fourth Edition [Measurement instrument]. Circle Pines, MN: American Guidance Services.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Retrieved from https://www.amazon.com/Essentials-Educational-Measurement-Robert-Ebel/dp/0132846136

Ekman, P. (2016). What scientists who study emotion agree about. *Perspectives on Psychological Science, 11*(1), 31–4. doi:10.1177/1745691615596992

Ekman, P., & Cordaro, D. (2011). What is meant by aalling emotions basic. *Emotion Review*, *3*(4), 364–370. doi:10.1177/1754073911410740

Ekman, P., Friesen, W. V, & Hager, J. (1978). *The Facial Action Coding System (FACS): A technique for the measurement of facial action* [Measurement Instrument]. Palo Alto, CA: Consulting Psychologists.

Ekman, P., Friesen, W. V, & Hager, J. C. (2002). *Facial Action Coding System* [Measurement Instrument]. Salt Lake City UT: Research Nexus division of Network Information Research Corporation.

Ekman P., & Friesen W.V. (1976). Pictures of Facial Affect. California, CA: Human Interaction Laboratory, University of California Medical Center.

Elfenbein, H. A., Foo, M. Der, White, J., Tan, H. H., & Aik, V. C. (2007). Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. *Journal of Nonverbal Behavior*, *31*(4), 205–223. doi:10.1007/s10919-007-0033-7

Ellison, J. W., Rosenfeld, J. A., & Shaffer, L. G. (2013). Genetic basis of intellectual disability. *Annual Review of Medicine*, *64*, 441–50. doi:10.1146/annurev-med-042711-140053

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572–595). Oxford, England: Oxford University Press.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low Kappa: The problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 543–549. doi:10.1016/0895-4356(90)90158-L

Field, A. P. (2005). *Discovering statistics using SPSS for Windows: Advanced techniques for beginners* (2nd ed.). London, England: Sage Publications.

Friesen, W, V., & Ekman, P. (1983). *EMFACS-7: Emotional facial action coding system* [Measurement Instrument]. San Francisco, CA: Unpublished manuscript, University of California, Human Interaction Laboratory.

Frijda, N. H., Kuipers, P., & ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology*, *57*(2), 212–228. doi:10.1037/0022-3514.57.2.212

Gagliardi, C., Frigerio, E., Burt, D. M., Cazzaniga, I., Perrett, D. I., & Borgatti, R. (2003). Facial expression recognition in Williams syndrome. *Neuropsychologia*, *41*(6), 733–8. doi: 10.1016/s0028-3932(02)00178-1

García-Villamisar, D., Rojahn, J., Zaja, R. H., & Jodra, M. (2010). Facial emotion processing and social adaptation in adults with and without autism spectrum disorder. *Research in Autism Spectrum Disorders* , *4*(755–762). doi:10.1016/j.rasd.2010.01.016

Gentsch, K., Grandjean, D., & Scherer, K. R. (2013). Temporal dynamics of event-related potentials related to goal conduciveness and power appraisals. *Psychophysiology*, *50*(10), 1010-1022. doi:10.1111/psyp.12079

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489. doi:10.5812/ijem.3505

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence*, *24(1)*, 13–23.

Grandjean, D., & Scherer, K. R. (2008). Unpacking the Cognitive Architecture of Emotion Processes. *Emotion*, *8*(3), 341–351. doi:10.1037/1528-3542.8.3.341

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24(2)*, 95–112. doi: 10.1007/BF02289823

Gwet, K. L. (2015). R functions for calculating agreement coefficients [Script file]. Retrieved from http://www.agreestat.com/r_functions.html

Gwet, K. L. (2008). Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, *61*(1), 29–48. doi:10.1348/000711006X126600

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (4th ed.). City of Gaithersburg, MD: Advanced Analytics, LLC.

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. *Trends in Cognitive Sciences*, *16*(9), 458–466. doi:10.1016/j.tics.2012.07.006

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322. doi:10.1007/s11065-010-9138-6

Harris, J. C. (2006). *Intellectual disability: Understanding its development, causes, classification, evaluation, and treatment*. Oxford, England: Oxford University Press. Retrieved from https://books.google.co.zw/books?id=VpARDAAAQBAJ

Harris, J. C., & Greenspan, S. (2016). Definition and nature of intellectual disability. In N. N. Singh (Ed.), *Handbook of evidence-based practices in Intellectual and Developmental Disabilities*. Cham, Switzerland: Springer International Publishing.

Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *American journal of mental retardation, 104*(3), 270–278. doi:10.1352/0895-8017(1999)104<0270:ROFEEF>2.0.CO;2

Heber, R. (1959). A manual on terminology and classification in mental retardation: A monograph supplement. *American Journal of Mental Deficiency*, *64(2)*, 1–111.

Hess, U., Blairy, S., & Kleck, R. E. (1997). The intensity of emotional facial expressions and decoding accuracy. *Journal of Nonverbal Behavior*, *21*(4), 241−257. doi: 10.1023/A:1024952730333

Hetzroni, O., & Oren, B. (2002). Effects of intelligence level and place of residence on the ability of individuals with mental retardation to identify facial expressions. *Research in Developmental Disabilities*, *23*(6), 369–378. doi:10.1016/S0891-4222(02)00139-7

Hoaken, P. N. S., Allaby, D. B., & Earle, J. (2007). Executive cognitive functioning and the recognition of facial expressions of emotion in incarcerated violent offenders, non-violent offenders, and controls. *Aggressive Behavior*, *33*(5), 412–421. doi:10.1002/ab.20194

Hobson, R. P. (1991). Methodological issues for experiments on autistic individuals' perception and understanding of emotion. *Journal of child psychology and psychiatry, and allied disciplines*, *32*(7), 1135–58. doi:10.1111/j.1469-7610.1991.tb00354.x

Hobson, R. P., Ouston, J., & Lee, A. (1989a). Recognition of emotion by mentally retarded adolescents and young adults. *American Journal of Mental Retardation*, *93*(4), 434–443.

Hobson, R. P., Ouston, J., & Lee, A. (1989b). Naming emotion in faces and voices: Abilities and disabilities in autism and mental retardation. *British Journal of Developmental Psychology*, *7*(3), 237–250. doi:10.1111/j.2044-835X.1989.tb00803.x

Hoffmann, H., Kessler, H., Eppel, T., Rukavina, S., & Traue, H. C. (2010). Expression intensity, gender and facial emotion recognition: Women recognize only subtle facial emotions better than men. *Acta psychologica*, *135*(3), 278–83. doi:10.1016/j.actpsy.2010.07.012

Howard, M.A, Cowell, P. E., Boucher, J., Broks, P., Mayes, A., Farrant, A., & Roberts, N. (2000). Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism. *Neuroreport*, *11*(13), 2931–2935. doi:10.1097/00001756-200009110-00020

Hwang, H., & Matsumoto, D. (2015). Evidence for the universality of facial expressions of emotion. In M. K. Mandal & A. Awasthi (Eds.), *Understanding Facial Expressions in Communication* (pp. 41–56). doi:10.1007/978-81-322-1934-7_3

Iarocci, G., & Petrill, S. A. (2012). Behavioral genetics, genomics, intelligence and mental retardation. In J. A. Burack, R. M. Hodapp, G. Iarocci, & E. Zigler (Eds.), *The Oxford handbook of intellectual disability and development* (pp. 12–29). New York, NY: Oxford University Press.

Izard. C. E. (1983). *The maximally discriminative facial movements coding system* Measurement instrument]. Delaware, DE: University of Delaware. Retrieved from https://www.library.yorku.ca/find/Record/1068572

Izard, C., Dougherty, L. M., & Hembree, E. A. (1983). *A system for identifying affect expressions by holistic judgments (AFFEX (*Revised ed.). Delaware, DE: University of Delaware.

Izard, C. E. (2010). The many meanings/aspects of emotion: definitions, functions, activation, and regulation. *Emotion Review*, *2*(4), 363–370. doi:10.1177/1754073910374661

Izard, C. E. (2011). Forms and functions of emotions: Matters of emotion–cognition interactions. *Emotion Review*, *3*(4), 371–378. doi:10.1177/1754073911410737

Izard, C. E., Fantauzzo, C. A., Castle, J. M., Haynes, O. M., Rayias, M. F., & Putnam, P. H. (1995). The ontogeny and significance of infants' facial expressions in the first 9 months of life. *Developmental Psychology*, *31*(6), 997–1013. doi:10.1037/0012-1649.31.6.997

Jensen, A. R. (2002). Psychometric g: Definition and substantiation. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 39–53). Retrieved from https://books.google.co.nz/books?id=lPN4AgAAQBAJ

Jones, K., & Day, J. D. (1997). Discrimination of two aspects of cognitive-social intelligence from academic intelligence. *Journal of Educational Psychology*, *89*(3), 486–497. doi:10.1037/0022-0663.89.3.486

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2001). Dynamic properties influence the perception of facial expressions. *Perception*, *30*(7), 875–887. doi:10.1068/p3131

Kaufman, L., Ayub, M., & Vincent, J. B. (2010). The genetic basis of non-syndromic intellectual disability: A review. *Journal of Neurodevelopmental Disorders*. doi:10.1007/s11689-010-9055-2

Kaulard, K., Cunningham, D. W., Bü Lthoff, H. H., Wallraven, C., & Ernst, M. O. (2012). The MPI Facial Expression Database — A validated database of emotional and conversational facial expressions. *PLoS ONE*, *7*(3), 1–18. doi:10.1371/journal.pone.0032321

Kendler, K. S., Halberstadt, L. J., Butera, F., Myers, J., Bouchard, T., Ekman, P. (2008). The similiarity of facial expressions in response to emotion-inducing films in reared-apart twins. *Psychological Medicine*, *38*(10), 1475–1483. doi:10.1017/S0033291707001535

Kennerknecht, I., Kischka, C., Stemper, C., Elze, T., & Stollhoff, R. (2011). Heritability of face recognition. In *Face Analysis, Modeling and Recognition Systems* (pp. 175–200). doi:10.5772/1837

Keselman, H. J., Rogan, J. C., Mendoza, J. L., & Breen, L. J. (1980). Testing the validity conditions of repeated measures F tests. *Psychological Bulletin*, *87*(3), 479–481. doi:10.1037/0033-2909.87.3.479

Kessler, H., Doyen-Waldecker, C., Hofer, C., Hoffmann, H., Traue, H. C., & Abler, B. (2011). Neural correlates of the perception of dynamic versus static facial expressions of emotion. GMS *Psycho-social medicine*, *8*, 1-8. doi:10.3205/psm000072

Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage*, *18*(1), 156–168. doi:10.1006/nimg.2002.1323

Koriakin, T. A., McCurdy, M. D., Papazoglou, A., Pritchard, A. E., Zabel, T. A., Mahone, E. M., & Jacobson, L. A. (2013). Classification of intellectual disability using the Wechsler Intelligence Scale for Children: Full Scale IQ or General Abilities Index? *Developmental Medicine and Child Neurology*, *55*(9), 840–5. doi:10.1111/dmcn.12201

Kozma, A., Mansell, J., & Beadle-Brown, J. (2009). Outcomes in different residential settings for people with intellectual disability: A systematic review. *American Journal On Intellectual and Developmental Disabilities*, *114*(3), 193–222. doi:10.1352/1944-7558-114.3.193

Kret, M. E. (2015). Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Frontiers in psychology*, *6*, 1–10. doi:10.3389/fpsyg.2015.00711

Lazarus, R. S. (1991). *Emotion and adaptation*. Retrieved from https://books.google.co.nz/books?id=9DxnDAAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Leppänen, J. M., & Hietanen, J. K. (2001). Emotion recognition and social adjustment in school-aged girls and boys. *Scandinavian Journal of Psychology*, *42*(5), 429–435. doi: 10.1111/1467-9450.00255

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198. doi:10.2307/2290157

Lowe, M. R., & Cautela, J. R. (1978). A self-report measure of social skill. *Behavior Therapy*, *9*(4), 535–544. doi:10.1016/S0005-7894(78)80126-9

MacMillan, D. L., Gresham, F. M., & Siperstein, G. N. (1993). Conceptual and psychometric concerns about the 1992 AAMR definition of mental retardation. *American journal of mental retardation, 98*(3), 325–35.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science*, *11*(3), 9–16. doi:10.1111/j.1467-7687.2008.00671.x

Marlowe, H. A. (1986). Social intelligence: Evidence for multidimensionality and construct independence. *Journal of Educational Psychology*, *78*(1), 52–58. doi:10.1037/0022-0663.78.1.52

Matson, J. L. (2009). *Social Performance Survey Schedule* [Measurement instrument]. Balton Rouge: LA: Disability Consultants.

Matson, J. L., & Hammer, D. (1996). Assessment of social functioning. In J. Jacobson, & A. Mulick (Eds.), *Manual of diagnosis and professional practice in mental retardation* (pp. 157–163). Washington, DC: American Psychological Association.

Matson, J. L., Helsel, W. J., Bellack, A. S., & Senatore, V. (1983). Development of a Rating Scale to Assess Social Skill Deficits in Mentally Retarded Adults. *Applied Research in Mental Retardation*, *4*(4), 399–407. doi:10.1016/0270-3092(83)90038-3

Matson, J. L., & Swiezy, N. (1994). Social skills training with autistic children. In J. L. Matson (Ed.), *Autism in children and adults: Etiology, assessment, and intervention*. Calafornia, CA: Brooks Cole.

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of Personality and Social Psychology*, *96*(1), 1–10. doi:10.1037/a0014037

Maulik, P. K., Mascarenhas, M. N., Mathers, C. D., Dua, T., & Saxena, S. (2011). Prevalence of intellectual disability: A meta-analysis of population-based studies. *Research in Developmental Disabilities*. doi:10.1016/j.ridd.2010.12.018

Mcalpine, C., Singh, N. N., Ellis, C. R., Kendall, K. A., & Hampton, C. (1992). Enhancing the ability of adults with mental retardation to recognize facial expressions of emotion. *Behavior Modification*, *16*(4), 559–573. doi:10.1177/01454455920164007

McKenzie, K., Matheson, E., McKaskie, K., Hamilton, L., & Murray, G. C. (2001). A picture of happiness. *Learning Disability Practice*, *4*(1), 26–29. doi:10.7748/ldp2001.05.4.1.26.c1451

McKenzie, K., Milton, M., Smith, G., & Ouellette-Kuntz, H. (2016). Systematic review of the prevalence and incidence of intellectual disabilities: Current trends and issues. *Current*

Developmental Disorders Reports, *3*(2), 104–115. doi:10.1007/s40474-016-0085-7

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630–633. doi:10.3758/BF03206543

Montagne, B., Kessels, R. P. C., De Haan, E. H. F., & Perrett, D. I. (2007). The Emotion Recognition Task: A paradigm to measure the perception of facial emotional expressions at different intensities. *Perceptual and motor skills*, *104*(2), 589–98. doi:10.2466/pms.104.2.589-598

Moore, D. G. (2000). Underestimating the emotion perception capacities of people with mental retardation. In E. Thommen & C. Vogel, (Eds.). In *Lire les passions* (pp. 79–96). Berne: Lang: Peter Lang Publishing.

Moore, D. G. (2001). Reassessing Emotion Recognition Performance in People With Mental Retardation: A Review. *American Journal on Mental Retardation*, *106*(6), 481. doi:10.1352/0895-8017(2001)106<0481:RERPIP>2.0.CO;2

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review*, *5*(2), 119–124. doi:10.1177/1754073912468165

Mortillaro, M., Meuleman, B., & Scherer, K. R. (2012). Advocating a componential appraisal model to guide emotion recognition. *International Journal of Synthetic Emotions*, *3*(1), 18–32. doi:10.4018/jse.2012010102

National Research Council (US) Committee on Disability Determination for Mental Retardation. (2002). The relationship of intelligence and adaptive behavior. In D. J. Reschly, T. G. Myers, & C. R. Hartel (Eds.), *Mental Retardation: Determining Eligibility for Social Security Benefits*. doi:10.17226/10295

NCS Pearson. (2012). *Raven's Standard Progressive Matrices (SPM v2.0) 2012 Norms Update Supplement*. Bloomington, MN: Author.

Nezlek, J. B., Vansteelandt, K., Van Mechelen, I., & Kuppens, P. (2008). Appraisal-emotion relationships in daily life. *Emotion*, *8*(1), 145–150. doi:10.1037/1528-3542.8.1.145

Nummenmaa, L., Glerean, E., Hari, R., & Hietanen, J. K. (2014). Bodily maps of emotions. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(2), 646–51. doi:10.1073/pnas.1321664111

O'Leary, U. M., Rusch, K. M., & Guastello, S. J. (1991). Estimating age-stratified WAIS-R IQs from scores on the Raven's Standard Progressive Matrices. *Journal of Clinical Psychology*, *47*(2), 277–284. doi: 10.1002/1097-4679(199103)47:2<277::AID-JCLP2270470215>3.0.CO;2-I

Owen, A., Browning, M., & Jones, R. S. P. (2001). Emotion recognition in adults with mild-moderate learning disabilities: An exploratory study. *Journal of Intellectual Disabilities*, *5*(3), 267–281. doi:10.1177/146900470100500309

Pallant, J. (2007). *SPSS survival manual : A step by step guide to data analysis using SPSS*. (3rd ed.). Sydney, Australia: Allen and Unwin.

Panksepp, J., & Watt, D. (2011). What is basic about basic emotions? lasting lessons from affective neuroscience. *Emotion Review*, *3*(4), 387–396. doi:10.1177/1754073911410741

Pochon, R., & Declercq, C. (2013). Emotion recognition by children with Down syndrome: a longitudinal study. *Journal of intellectual & developmental disability*, *38*(4), 332–43. doi:10.3109/13668250.2013.826346

Poljac, E., Poljac, E., & Wagemans, J. (2012). Reduced accuracy and sensitivity in the perception of emotional facial expressions in individuals with high autism spectrum traits. *Autism*, *17*(6), 668–680. doi:10.1177/1362361312455703

Pollak, S. D., & Kistler, D. J. (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(13), 9072–9076. doi:10.1073/pnas.142165999

Power, M. J., & Dalgleish, T. (2016). *Cognition and emotion: From order to disorder* (3rd ed). Retrieved from https://books.google.co.nz/books?id=ILFmCgAAQBAJ&printsec =frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Qualitrics Survey Software (Version 37 892) [Computer software]. (2013). Provo, UT: Retrieved from www.qualtrics.com

Raven, J., C, R. J., & Court, J. H. (2000). *Raven's Manual Section 3 - Standard Progressive Matrices 1998 Edition*. Oxford, England: Oxford Psychologists Press.

Robinson, L., Spencer, M. D., Thomson, L. D. G., Sprengelmeyer, R., Owens, D. G. C., Stanfield, A. C., … Johnstone, E. C. (2012). Facial emotion recognition in Scottish prisoners. *International Journal of Law and Psychiatry*, *35*, 57–61. doi:10.1016/j.ijlp.2011.11.009

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition: Interpretive Manual*. Itasca, IL: Riverside Publishing.

Rojahn, J., Esbensen, A., & Hoch, T. (2006). Relationships between facial discrimination and social adjustment in mental retardation. *American Journal of Mental Retardation*, *111*(5), 366–377. doi:10.1352/0895-8017(2006)111[366:RBFDAS]2.0.CO;2

Rojahn, J., Lederer, M., & Tassé, M. J. (1995). Facial emotion recognition by persons with mental retardation: A review of the experimental literature. *Research in Developmental Disabilities*, *16*(5), 393–414. doi:0891-4222(95)00019-J [pii]

Rojahn, J., Rabold, D. E., & Schenider, F. (1995). Emotion specificity in mental retardation. *American Journal on Mental Retardation*, *99*(5), 477–486.

Rojahn, J., & Warren, V. J. (1997). Emotion recognition as a function of social competence and depressed mood in individuals with intellectual disability. *Journal of Intellectual Disability Research*, *41*(6), 469–475. doi:10.1111/j.1365-2788.1997.tb00738.x

Rojahn, J., & Zaja, R. H. (2007). The emotion specificity hypothesis in intellectual disabilities. *Psychology in Intellectual and Developmental Disabilities - Official Publication of Division 33(2)*, 4-6.

Roseman, I. J., Wiest, C., & Swartz, T. S. (1994). Phenomenology, behaviors, and goals differentiate discrete emotions. *Journal of Personality and Social Psychology*, *67*(2), 206–221. doi:10.1037/0022-3514.67.2.206

Rosenberg, E. L., & Ekman, P. (1995). Conceptual and methodological issues in the judgment of facial expressions of emotion. *Motivation and Emotion*, *19*(2), 111–138. doi:10.1007/BF02250566

Rosenthal, R. (1994). Parametric measures of effect size. In L. V. Hedges & H. M. Cooper (Eds.), *The Handbook of Research Synthesis* (pp. 231–244). New York, NY: Russell Sage Foundation.

Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: A bivariate boxplot. *The American Statistician*, *53*(4), 382–387. doi:10.1080/00031305.1999.10474494

Roy, M., Retzer, A., & Sikabofori, T. (2015). Personality development and intellectual disability. *Current Opinion in Psychiatry*, *28*(1), 35–39. doi:10.1097/YCO.0000000000000118

RStudio: Integrated Development for R (Version 0.99.902) [Computer Program]. (2016). Boston, MA. Retrieved from www.rstudio.com.

Rymarczyk, K., Biele, C., Grabowska, A., & Majczynski, H. (2011). EMG activity in response to static and dynamic facial expressions. *International Journal of Psychophysiology*, *79*(2), 330–333. doi:10.1016/j.ijpsycho.2010.11.001

Sappok, T., Gaul, I., Bergmann, T., Dziobek, I., Bölte, S., Diefenbacher, A., & Heinrich, M. (2014). The Diagnostic Behavioral Assessment for Autism Spectrum Disorder - Revised: A screening instrument for adults with intellectual disability suspected of autism spectrum disorders. *Research in Autism Spectrum Disorders*, *8*(4), 362–375. doi:10.1016/j.rasd.2013.12.016

Sauter, D. A., LeGuen, O., & Haun, D. B. M. (2011). Categorical perception of emotional facial expressions does not require lexical categories. *Emotion, 11*(6), 1479–1483. doi:10.1037/a0025336

Schalock, R. L., Borthwick-Duffy, S. A., Bradley, V. J., Buntinx, W. H., Coulter, D. L., Craig, E. M., … Shogren, K. A. (2010). *Intellectual disability: Definition, classification, and systems of supports* (11th ed.). Washington, DC: American Association on Intellectual and Developmental Disabilities.

Schalock, R. L., Luckasson, R., J, B. V., Buntinx, W. H. E., Lachapelle, Y., Shogren, K. A., … Wehmeyer, M. L. (2012). *User's Guide to Intellectual Disability: Definition, Classification, and Systems of Supports* (11th ed). Washington, DC: American Association on Intellectual and Developmental Disabilities.

Scheffer, J. (2002). Dealing with missing data. *Research letters in the information and mathematical sciences*, *3*, 153–160. Retrieved from http://mro.massey.ac.nz/xmlui/handle/10179/4332

Scherer, K. R. (1987). Toward a dynamic theory of emotion. *Geneva Studies in Emotion*, *1*, 1–96.

Scherer, K. R. (1994). Toward a concept of "modal emotions." In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 25–31). Oxford, England: Oxford University Press.

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, *7*(1), 113–130. doi:10.1037/1528-3542.7.1.113

Schlosberg, H., & Harold. (1954). Three dimensions of emotion. *Psychological Review*, *61*(2), 81–88. doi.org/10.1037/h0054570

Scotland, J. L., Cossar, J., & McKenzie, K. (2015). The ability of adults with an intellectual disability to recognise facial expressions of emotion in comparison with typically developing individuals: A systematic review. *Research in developmental disabilities*, *41–42*, 22–39. doi:10.1016/j.ridd.2015.05.007

Scotland, J. L., McKenzie, K., Cossar, J., Murray, A., & Michie, A. (2016). Recognition of facial expressions of emotion by adults with intellectual disability: Is there evidence for the emotion specificity hypothesis? *Research in developmental disabilities*, *48*, 69–78. doi:10.1016/j.ridd.2015.10.018

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals* (Fourth ed.) [Measurement intrument]. San Antonio: TX: Psychological Corporation.

Simon, E. W., Rosen, M., Grossman, E., & Pratowski, E. (1995). The relationships among facial emotion recognition, social skills, and quality of life. *Research in Developmental Disabilities*, *16*(5), 383–391. doi:10.1016/0891-4222(95)00025-I

Smith, C. A., & Scott, H. S. (1997). A componential approach to the meaning of facial expression. In J. A. Russel & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 229–254). Cambridge, England: Cambridge University Press.

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (2005). *Vineland Adaptive Behavior Scales, Second Edition (Vineland-II)* [Measurement instrument]. Circle Pines, MN: Pearson Assessments.

Sparrow, S. S., Balla, D. A., Cicchetti, D. V, Harrison, P. L., & Doll, E. A. (1984). *Vineland Adaptive Behavior Scales, Interview Edition, Survey Form manual*. Circle Pines, MN: American Guidance Service Publishing.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (Fourth ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Stevenson, A. (Ed.). (2010). *Oxford dictionary of English* (3rd ed.). doi:10.1093/acref/9780199571123.001.0001

SurveyMonkey (2015). FluidSurveys [software]. Retrieved from http://fluidsurveys.com

Switzky, H., & Greenspan, S. (2006). *What is mental retardation? Ideas for an evolving disability in the 21st century*. Washington, DC: American Association on Mental Retardation.

Tassé, M. (2009). Adaptive behavior assessment and the diagnosis of mental retardation in capital cases. *Applied Neuropsychology*, *16*, 114–123. doi:10.1080/09084280902864451

Thompson, B., & Levitov, J. E. (1985). Using Microcomputers to Score and Evaluate Items. *Collegiate Microcomputer*, *3*(2), 163–68.

Thorndike, R. M., & Thorndike-Christ, T. M. (2009). *Measurement and Evaluation in Psychology and Education* (Eigth ed.). Essex, England: Pearson Education.

Tian, Y., Kanade, T., & Cohn, J. F. (2005). Facial Expression Analysis. In S. Z. Li & A. K. Jain (Eds.), *Handbook of Face Recognition* (pp. 247–275). New York: Springer-Verlag. doi:10.1007/b138828

Tomkins, S. S. (Ed.). (1962). *Affect, imagery, consciousness: Volume 1 - The positive affects*. New York, NY: Springer.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology*, *17*(1–3), 257–279. doi:10.1080/026432900380607

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., … Nelson, C.

(2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249. doi:10.1016/j.psychres.2008.05.006

Tukey, J. (1974). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians*, *2*, 523–532.

Turk, J., Cornish, K., & Cornish, K. (1998). Face recognition and emotion perception in boys with Fragile-X syndrome. *Journal of Intellectual Disability Research*, *42*(6), 490–499. doi:10.1046/j.1365-2788.1998.4260490.x

Uljarevic, M., & Hamilton, A. (2013). Recognition of emotions in autism: A formal meta-analysis. *Journal of Autism and Developmental Disorders*, *43*(7), 1517–1526. doi:10.1007/s10803-012-1695-5

Hambleton, R. K., & Dirir, M. (2003). Classical and modern item analysis. In R. Fernandez-Ballesteros (Ed.). *Encyclopedia of Psychological Assessment* (pp. 188–191). Retrieved from https://play.google.com/store/books/details?id=UDzXlxE6jEwC&rdid=book-UDzXlxE6jEwC&rdot=1&source=gbs_vpt_read&pcampaignid=books_booksearch_viewport

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale: Technical and interpretive manual.* San Antonio: TX: Pearson.

Weinfurt, K. P. (2000). Repeated measures analyses: ANOVA, MANOVA, and HLM. In L. G. Grimm & P. R. Yarnold (Ed.), *Reading and understanding more multivariate statistics* (pp. 317-361). Washington, DC: American Psychological Association.

Weis, R. (2013). *Introduction to Abnormal Child and Adolescent Psychology* (2nd ed). Retrieved from https://books.google.co.nz/books?id

Weiss, B., Weisz, J. R., & Bromfield, R., (1986). Performance of retarded and nonretarded persons on information- processing tasks: Further tests of the similar structure hypothesis. *Psychological Bulletin*, *100*(2), 157–175. doi: 10.1037/0033-2909.100.2.157

Weisz, J. R. (1999). Cognitive performance and learned helplessness in mentally retarded persons. In E. Zigler & D. Bennett-Gates (Eds.), *Personality development in individuals with mental* *retardation* (pp. 17–46). Cambridge, England: Cambridge University Press.

Weisz, J. R., & Yeates, K. O. (1981). Cognitive Development in Retarded and Nonretarded Persons: Piagetian Tests of the Similar Structure Hypothesis. *Psychological Bulletin*, *90*, 153–118. doi: 10.1037/0033-2909.90.1.153

Weisz, J. R., & Zigler, E. (1979). Cognitive Development in retarded and nonretarded persons: Piagetian tests of the similar sequence hypothesis. *Psychological Bulletin*, *86*(4), 831–851. doi: 10.1037/0033-2909.86.4.831

Wessa, P. (2012). Bagplot (v1.0.1) in Free Statistics Software (Version1.1.23-r7) [Software]. Retrieved from www.wessa.net/rwasp_bagplot.wasp

Whittington, J., & Holland, T. (2011). Recognition of emotion in facial expression by people with Prader-Willi syndrome. *Journal of Intellectual Disability Research*, *55*(1), 75–84. doi:10.1111/j.1365-2788.2010.01348.x

Williams, K. R., Wishart, J. G., Pitcairn, T. K., & Willis, D. S. (2005). Emotion recognition by children With Down Syndrome: Investigation of specific impairments and error patterns. *American Journal on Mental Retardation*, *110*(5), 378. doi:10.1352/0895-8017(2005)110[378:ERBCWD]2.0.CO;2

Williams, K. T. (2007). *Expressive Vocabulary Test, Second Edition* [Measurement instrument]. Circle Pines, MN: American Guidance Service Publishing.

Wilson, N. J., Jaques, H., Johnson, A., & Brotherton, M. L. (2016). From social exclusion to supported inclusion: adults with intellectual disability discuss their lived experiences of a structured social group. *Journal of Applied Research in Intellectual Disabilities*. Advance online publication. doi:10.1111/jar.12275

Wingenbach, T. S. H., Ashwin, C., & Brosnan, M. (2016). Validation of the Amsterdam Dynamic Facial Expression Set - Bath Intensity Variations (ADFES-BIV): A set of videos expressing low, intermediate, and high intensity emotions. *PLoS ONE*, *11*(1), 1–28. doi:10.1371/journal.pone.0147112

Wishart, J. G., Cebula, K. R., Willis, D. S., & Pitcairn, T. K. (2007). Understanding of facial expressions of emotion by children with intellectual disabilities of differing aetiology. *Journal of Intellectual Disability Research*, *51*(7), 551–563. doi:10.1111/j.1365-2788.2006.00947.x

Wood, D. A. (1974). *Test Construction: Development and Interpretation of Achievement Tests* (2nd ed.). Columbus, OH: Charles E. Merrill Publishing Company.

Wood, P. M., & Kroese, B. S. (2007). Enhancing the emotion recognition skills of individuals with learning disabilities: A review of the literature. *Journal of Applied Research in Intellectual Disabilities*, *20*(6), 576–579. doi:10.1111/j.1468-3148.2006.00355.x

World Health Organization. (1992). *Assessment of people with mental retardation*. Retrieved from https://extranet.who.int/iris/restricted/bitstream/10665/62321/1/WHO_MNH_PSF_92.3.pdf

World Health Organization. (1996). *ICD-10 Guide for Mental Retardation; Division of Mental Health and Prevention of Substance Abuse*. Geneva, Switzerland: Author.

World Health Organization. (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva, Switzerland: Author.

Yager, J. A., & Ehmann, T. S. (2006). Untangling social function and social cognition: a review of concepts and measurement. *Psychiatry: Interpersonal & Biological Processes*, *69*(1), 47–68. doi: 10.1521/psyc.2006.69.1.47

Yoshikawa, S., & Sato, W. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, *104*, 1–18. doi:10.1016/j.cognition.2006.05.001

Young, A. W., Rowland, D., Calder, A. J., Etcoff, N. L., Seth, A., & Perrett, D. I. (1997). Facial expression megamix: tests of dimensional and category accounts of emotion recognition. *Cognition*, *63*(3), 271–313. doi: 10.1016/S0010-0277(97)00003-6

Zaja, R. H., & Rojahn, J. (2008). Facial emotion recognition in intellectual disabilities. *Current Opinion in Psychiatry*, *21*(5), 441–444. doi:10.1097/YCO.0b013e328305e5fd

Zak, M., Laeng, B., & Simon-Liedtke, J. T. (2015). Can the recognition of emotional expressions be enhanced by manipulating facial skin color? *Colour and Visual Computing Symposium (CVCS),* 1–6. doi:10.1109/CVCS.2015.7274897

Zigler, E. (1969). Developmental versus difference theories of mental retardation and the problem of motivation. *American Journal of Mental Deficiency*, *73*(4), 536–556.

Zigler, E., Bennett-Gates, D., Hodapp, R., & Henrich, C. C. (2002). *Assessing personality traits of individuals with mental retardation. American Journal on Mental Retardation*, *107(3),* 181-193. doi:10.1352/0895-8017(2002)107<0181:APTOIW>2.0.CO;2

APPENDIX A

DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS-FIFTH EDITION CRITERIA FOR INTELLECTUAL DISABILITY

Intellectual disability (intellectual developmental disorder) is a disorder with onset during the developmental period that includes both intellectual and adaptive functioning deficits in conceptual, social, and practical domains. The following three criteria must be met:

A. Deficits in intellectual functions, such as reasoning, problem solving, planning, abstract thinking, judgment, academic learning, and learning from experience, confirmed by both clinical assessment and individualized, standardized intelligence testing.
B. Deficits in adaptive functioning that result in failure to meet developmental and sociocultural standards for personal independence and social responsibility. Without ongoing support, the adaptive deficits limit functioning in one or more activities of daily life, such as communication, social participation, and independent living, across multiple environments, such as home, school, work, and community.
C. Onset of intellectual and adaptive deficits during the developmental period.

(DSM-5; American Psychiatric Association, 2013, p. 71)

APPENDIX B

GWET AC$_1$ SCRIPT FILE

```
                              AGREE.COEFF3.DIST.R
#                          (September 26, 2015)
#Description: This script file contains a series of R functions for computing various
 agreement coefficients
#           for multiple raters (2 or more) when the input data file is in the form of
 nxq matrix or data frame showing
#            the count of raters by subject and by category. That is n = number of
 subjects, and q = number of categories.
#            A typical table entry (i,k) represents the number of raters who classified
 subject i into category k.
#Author: Kilem L. Gwet, Ph.D. (Please send comments to: gwet@agreestat.com. Thank you)
#

#==========================================================================================
 ====
#gwet.ac1.dist: Gwet's AC1/Ac2 coefficient (Gwet(2008)) and its standard error for
 multiple raters when input
#            dataset is a nxq matrix representing the distribution of raters by subject
 and by category.
#-------------
#The input data "ratings" is an nxq matrix showing the number of raters by subject and
 category. A typical entry associated
#with a subject and a category, represents the number of raters who classified the
 subject into the specified category. Exclude
#all subjects that are not rated by any rater.
#Bibliography:
#Gwet, K. L. (2008). ``Computing inter-rater reliability and its variance in the
 presence of high
#         agreement." British Journal of Mathematical and Statistical Psychology, 61,
 29-48.
#==========================================================================================
gwet.ac1.dist <- function(ratings,weights="unweighted",conflev=0.95,N=Inf,print=TRUE){
  agree.mat <- as.matrix(ratings)
  n <- nrow(agree.mat) # number of subjects
  q <- ncol(agree.mat) # number of categories
  f <- n/N # final population correction

  # creating the weights matrix

  if (is.character(weights)){
     weights.mat<-diag(q)
  }else weights.mat= as.matrix(weights)

  agree.mat.w <- t(weights.mat%*%t(agree.mat))

  # calculating gwet's ac1 coefficient

  ri.vec <- agree.mat%*%rep(1,q)
  sum.q <- (agree.mat*(agree.mat.w-1))%*%rep(1,q)
  n2more <- sum(ri.vec>=2)
  pa <- sum(sum.q[ri.vec>=2]/((ri.vec*(ri.vec-1))[ri.vec>=2]))/n2more

  pi.vec <- t(t(rep(1/n,n))%*%(agree.mat/(ri.vec%*%t(rep(1,q)))))
  pe <- sum(weights.mat) * sum(pi.vec*(1-pi.vec)) / (q*(q-1))
  gwet.ac1 <- (pa-pe)/(1-pe)

  # calculating variance, stderr & p-value of gwet's ac1 coefficient
```

```
  den.ivec <- ri.vec*(ri.vec-1)
  den.ivec <- den.ivec - (den.ivec==0) # this operation replaces each 0 value with -1 to
make the next ratio calculation always possible.
  pa.ivec <- sum.q/den.ivec

  pe.r2 <- pe*(ri.vec>=2)
  ac1.ivec <- (n/n2more)*(pa.ivec-pe.r2)/(1-pe)
  pe.ivec <- (sum(weights.mat)/(q*(q-1))) * (agree.mat%*%(1-pi.vec))/ri.vec
  ac1.ivec.x <- ac1.ivec - 2*(1-gwet.ac1) * (pe.ivec-pe)/(1-pe)

  var.ac1 <- ((1-f)/(n*(n-1))) * sum((ac1.ivec.x - gwet.ac1)^2)
  stderr <- sqrt(var.ac1)# ac1's standard error
  p.value <- 2*(1-pt(gwet.ac1/stderr,n-1))

  lcb <- gwet.ac1 - stderr*qt(1-(1-conflev)/2,n-1) # lower confidence bound
  ucb <- min(1,gwet.ac1 + stderr*qt(1-(1-conflev)/2,n-1)) # upper confidence bound
  if(print==TRUE) {
    if (!is.character(weights)) {
   cat("Gwet's AC2 Coefficient\n")
        cat('==========================\n')
   cat('AC2 coefficient:',gwet.ac1,'\n')
        cat('Standard error:',stderr,'\n')
        cat(conflev*100,'% Confidence Interval: (',lcb,',',ucb,')\n')
        cat('P-value: ',p.value,'\n')
   cat('Percent agreement:',pa,'\n')
   cat('Percent chance agreement:',pe,'\n')
   cat('Weights:\n')
   cat('********\n')
   write.table(weights,row.names=FALSE,col.names=FALSE)
        cat('\n')
    }else{
   cat("Gwet's AC1 Coefficient\n")
        cat('==========================\n')
   cat('AC1 coefficient:',gwet.ac1,'\n')
   cat('Standard error:',stderr,'\n')
        cat(conflev*100,'% Confidence Interval: (',lcb,',',ucb,')\n')
        cat('P-value: ',p.value,'\n')
    }
  }
  invisible(c(pa,pe,gwet.ac1,stderr,p.value))
}
```

(Gwet, 2015)

APPENDIX C

FACS ACTION UNIT DESCRIPTORS AND UNDERLYING FACIAL MUSCULATURE

AU	Description	Muscular Basis
1	Inner Brow Raiser	Frontalis, pars medialis
2	Outer Brow Raiser	Frontalis, pars lateralis
4	Brow Lowerer	Corrugator supercilii, Depressor supercilii
5	Upper Lid Raiser	Levator palpebrae superioris
6	Cheek Raiser	Orbicularis oculi, pars orbitalis
7	Lid Tightener	Orbicularis oculi, pars palpebralis
9	Nose Wrinkler	Levator labii superioris alaquae nasi
10	Upper Lip Raiser	Levator labii superioris
11	Nasolabial Deepener	Zygomaticus minor
12	Lip Corner Puller	Zygomaticus major
13	Cheek Puffer	Levator anguli oris (a.k.a. Caninus)
14	Dimpler	Buccinator
15	Lip Corner Depressor	Depressor anguli oris (a.k.a. Triangularis)
16	Lower Lip Depressor	Depressor labii inferioris
17	Chin Raiser	Mentalis
18	Lip Puckerer	Incisivii labii superioris and Incisivii labii inferioris
20	Lip stretcher	Risorius w/ platysma
22	Lip Funneler	Orbicularis oris
23	Lip Tightener	Orbicularis oris
24	Lip Pressor	Orbicularis oris
25	Lips part**	Depressor labii inferioris or relaxation of Mentalis, or
26	Jaw Drop	Masseter, relaxed Temporalis and internal Pterygoid
27	Mouth Stretch	Pterygoids, Digastric
28	Lip Suck	Orbicularis oris
41	Lid droop	Relaxation of Levator palpebrae superioris
42	Slit	Orbicularis oculi
43	Eyes Closed	Relaxation of Levator palpebrae superioris; Orbicularis oculi,
44	Squint	Orbicularis oculi, pars palpebralis
45	Blink	Relaxation of Levator palpebrae superioris; Orbicularis oculi,
46	Wink	Relaxation of Levator palpebrae superioris; Orbicularis oculi,

(Ekman et al., 2002)

APPENDIX D

PRELIMINARY ITEM LEVEL INTER-RATER RELIABILITY ESTIMATES BASED ON NINE INDEPENDENT REVIEWERS

Subscale/Item	Percent correct	AC_1	Naturalness (%)
Anger			
Item 1	100	1.00	57.1
Item 2	88.9	.77	64.4
Item 3	88.9	.77	77.4
Item 4	88.9	.77	57.8
Disgust			
Item 5	88.9	.77	66.0
Item 6	100	1.00	57.2
Item 7	88.9	.77	72.3
Item 8	100	1.00	67.5
Fear			
Item 9	77.8	.59	51.7
Item 10	100	1.00	61.3
Item 11	70.0	.46	55.2
Item 12	77.8	.59	65.9
Happiness			
Item 13	100	1.0	74.3
Item 14	88.9	.77	89.3
Item 15	88.9	.77	71.9
Item 16	88.9	.77	69.8
Sadness			
Item 17	100	1.00	75.9
Item 18	88.9	.77	78.9
Item 19	100	1.00	77.3
Item 20	88.9	.77	76.6
Surprise			
Item 21	100	1.00	58.5
Item 22	100	1.00	59.5
Item 23	100	1.00	54.9
Item 24	88.9	.77	60.4

APPENDIX E-1

INTELLECTUAL DISABILITY SERVICES RECRUITMENT LETTER

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

Research project overview

Thank you for showing an interest in this project. Please read this information sheet carefully before deciding whether or not to nominate your facility/service to participate. If you decide to participate, we thank you. If you decide not to take part, there will be no disadvantage to you or your facility/service and we thank you for considering our request.

Project Summary

People with intellectual disability often experience impairments in social and vocational adjustment, as well as the development of psychopathology in addition to their social difficulties. It remains unclear whether reading facial emotions is a main contributor to the social difficulties experienced by this group.

The purpose of this study is to trial a computer based tool designed to test the emotion recognition abilities of people with an intellectual disability. The study will consider how cues such as movement and emotion intensity affect emotion recognition performance, and will also explore the link between emotion recognition and social functioning. We hope that this research will help researchers clarify the link between emotion recognition and social skills, and in time help coordinate better care and social learning opportunities for individuals with an intellectual disability.

The coordinating researcher on this project is Zara Godinovich who is completing this study in partial fulfilment of the degree Doctor of Clinical Psychology. The supervisors for this project are Dr Ian de Terte, and Dr Richard Fletcher, both senior lecturers for the School of Psychology, Massey University.

Participants and Recruitment

We are seeking up to 40 individuals with a diagnosis of mild to moderate intellectual disability. An equal number of men and women are required.

Participants must:

- be18 years or older;
- have received a diagnosis of intellectual disability as per the Diagnostic and Statistical Manual of Mental Disorders (Fourth or Fifth edition);
- not have been diagnosed with autism spectrum disorder.

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

It is preferable that participants:

- have a Full Scale IQ between 50 and 70 as established by the Wechsler Adult Intelligence Scales Fourth edition (WAIS-IV) (please let us know if this criterion severely limits the number of potential participants from your service);
- have non-syndromic intellectual disability (i.e. no neurological or genetic abnormality, no history of traumatic head injury, no serious psychiatric disorder, and no family history of intellectual disability).

Each participant will need to have member of support staff who is willing to provide information about the participant's social functioning. This person should be very familiar to the participant (i.e. has known the individual for at least six months). For participants without a legal Welfare Guardian written consent must be offered by a family member or the Area manager. However should a participant be under the care of a Welfare Guardian consent must be obtained from him or her.

Participants will be offered a $10.00 NZD Warehouse voucher as a token of our appreciation.

Project and Procedures

Participants based in Auckland may choose to meet the researcher at the Massey University Centre for Psychology based in Albany North Shore City. Alternatively the researcher will travel to your facility.

Individuals with intellectual disability
Participants will be asked to give written consent before completing a series of computerised tasks. The tasks will mostly involve discriminating between facial emotions, in addition to naming the colours of shapes and identifying individual's ages. The participants will also complete a test of receptive language. It is expected that the entire session will take about two hours and participants may request a break at anytime.

Support staff
The support or care staff member will complete two measures of social functioning and an autism screening measure. It is preferable that the support staff member be present during at least some of the testing session to help the participant feel more comfortable. Participants may choose to have a family member attend the session.

It is important that you are aware that one of the tests will screen for autistic traits. We would like to stress that these are tests used for research purposes and do not offer the same information as a comprehensive clinical assessment. Should a participant score outside the "normal range" this does not necessary equate to any abnormality or problem. For this reason we do not intend to share these scores with anyone outside of the research team. We will however give the legal guardians of the participants the option to have their dependent's GP notified should the supervising Clinical Psychologist identify any potential concerns.

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Participant Rights

Your service is under no obligation to accept this invitation. Both the service and participating individuals from your service have the right to:

- decline to answer any particular question;
- withdraw from the study anytime up to two weeks following the testing session;
- ask any questions about the study at any time during participation;
- provide information on the understanding that their name will not be used unless they give permission to the researcher;
- request a summary of the project findings when it is concluded.

Data Management

The results of the project may be published and will be made available in the lead researcher's doctoral thesis. Confidentiality will be maintained. We will not share any participant's personal scores with anyone outside the research team unless they are in a deidentified form.

Ethical Approval

This project has been reviewed and approved by Northern A Health and Disability Ethics Committee (HDEC: ref 14/NTA/64).

Project Contacts

If you would like to nominate your service/facility to take part in the study please contact Zara Godinovich. If you have any questions or would like further information you are welcome to contact either myself or my supervisors:

Zara Godinovich
Doctorate of Clinical Psychology student
Massey University
zara.godinovich@outlook.com
021 1694768

Dr Ian de Terte
Senior Lecturer and Clinical Psychologist
Massey University
i.deterte@massey.ac.nz
+64 4 801 5799, Ext 62033

Dr Richard Fletcher
Senior Lecturer
Massey University
r.b.fletcher@massey.ac.nz
+64 9 414-0800, Ext 41213

Thank you for reading this information sheet.

Yours sincerely,

Zara Godinovich

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

APPENDIX E-2

INFORMATION AND CONSENT FORM FOR PARTICIPANTS WITH INTELLECTUAL DISABILITY

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

Participant information summary and consent form

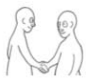
This study is to learn more about how people with an intellectual disability recognise emotions on other people's faces.

You will be shown pictures and videos of people's faces on a computer. You will be asked to guess how they are feeling or to guess how old they are. You will also be shown some moving shapes and will be asked what colour they are.

You will also do an activity where the researcher will show you four pictures and then read out a word. You will need to point to the picture that matches the word.

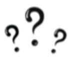
I will also ask your care manager some questions about you and about the different assessments you have done in the past. This will include health information, information about previous diagnoses, and past intelligence testing.

I will write about what we find out from working with you. This may be published in a scientific journal, which is like a book. We will not put your name in the journal so no one will know who you are.



We will ask your care manager or caregiver some questions about how you interact with other people.

If you get tired you can ask for a break.



If you have any questions you can ask me at anytime.

If you need to complain about anything, please tell me or talk to your care manager.

You will get a $10 voucher for taking part.

It is your decision to take part

You can say "Yes" 

You can say "No" 

If you understand what I have told you and you want to be involved in the study please sign your name.

Participant ____ Date __/__/__

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

APPENDIX E-3

THIRD PARTY CONSENT FORM FOR PARTICIPANTS WITH INTELLECTUAL DISABILITY

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

Third party consent form

I have read and understood the Letter of Invitation concerning this project. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage. I know that:

1. Participation by [redacted] clientele is entirely voluntary;
2. Only participants who consent will participate in the project;
3. Participants may withdraw from the study at anytime without any disadvantage;
4. Personal identifying information will be destroyed at the conclusion of the project but any raw data on which the results of the project depend will be retained in secure storage for five years, after which they will be destroyed;
5. This project should not cause any discomfort or risks;
6. No formal payment will be offered, however the participant will receive a Warehouse voucher for the value of $10.00 as a token of appreciation for their participation;
7. The results of the project may be published and may be made available on the internet but every attempt will be made to preserve anonymity.

Would you like your client's GP to be notified if the supervising Clinical Psychologist identifies any potential concerns relating to the autism screening measure? If yes, please list the GP's details bon the reverse

☐ Yes ☐ No

I give consent for you to approach the following Idea Service clientele to participate in the project "Assessing Facial Affect Recognition in People with Intellectual Disability" .

1. ______________________________ 4. ______________________________
2. ______________________________ 5. ______________________________
3. ______________________________ 6. ______________________________

Full Name:

Signature: ________________________________ **Date:** _____/_____/ 2014

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

APPENDIX E-4

INFORMATION SHEET FOR CHRONOLOGICALLY AGE MATCHED PARTICIPANTS

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

Participant information sheet

Thank you for showing an interest in this project. Please read this information sheet carefully before deciding whether or not to participate. If you decide to participate, we thank you. If you decide not to take part, there will be no disadvantage to you of any kind and we thank you for considering our request.

Project summary

People with intellectual disability often experience impairments in social and vocational adjustment, as well as the development of psychopathology in addition to their social difficulties. It remains unclear whether reading facial emotions is a main contributor to the social difficulties experienced by this group.

The purpose of this study is to trial a computer based tool designed to test the emotion recognition abilities of people with an intellectual disability. The study will consider how cues such as movement and emotion intensity affect emotion recognition performance, and will also explore the link between emotion recognition and social functioning.

We would like to use this tool to better understand how people with an intellectual disability perform on emotion recognition tasks. To achieve this we need to establish how individuals without intellectual disability would perform on such a task and it is for this reason that we invite you to participate. We hope that this research will help researchers clarify the link between emotion recognition and social skills and in time help coordinate better care and learning opportunities for individuals with an intellectual disability.

The coordinating researcher on this project is Zara Godinovich who is completing this study in partial fulfillment of the degree Doctor of Clinical Psychology. The supervisors for this project are Dr Richard Fletcher and Dr Ian de Terte, both senior lecturers for the School of Psychology, Massey University.

What type of participant is being sought?

This research study seeks participants who fit the age, ethnicity and gender profiles listed at the end of this information sheet. Unfortunately, we cannot include individuals with an autism spectrum diagnosis or an intellectual disability.

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

What will participants be asked to do?

Should you agree to take part in this project, you will be asked to complete a series of computerized tasks, most of which will involve identifying facial emotions. You will also complete a pattern completion task, and a questionnaire about how you interact with others.

It is important that you are aware that some of these tests will screen for autistic traits and be used to *infer* your IQ (intelligence quotient). We would like to stress that these are tests used for research purposes and do not offer the same information as a comprehensive clinical assessment. Should you score outside the 'normal range' this does not necessary equate to any abnormality or problem. For this reason we do not intend to share your scores with anyone outside of the research team. However, if you do have concerns you are welcome to list the contact details for your GP on the attached consent form, and we will pass on the results should the supervising Clinical Psychologist identify any potential concerns.

It is expected that the entire session will take about two hours. You may request a break at anytime. Sessions will be held at the Massey University Psychology Centre in Albany Village North Shore City. Should travel interfere with you ability to participate please contact the lead researcher.

Participant rights

You are under no obligation to accept this invitation. If you decide to participate, you have the right to:

- decline to answer any particular question;
- withdraw from the study anytime up to two weeks post your testing session;
- ask any questions about the study at any time during participation;
- provide information on the understanding that your name will not be used unless you give permission to the researcher;
- request a summary of the project findings when it is concluded.

Data management

The results of the project may be published and will be made available in the lead researcher's doctoral thesis. Confidentiality will be maintained. We will not share any of your personal scores with anyone outside the research team unless they are in a deidentified form.

Ethical Approval

This project has been reviewed and approved by Northern A Health and Disability Ethics Committee (HDEC: ref 14/NTA/64).

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Project Contacts

If you would like to take part in the study please contact Zara Godinovich. If you have any questions or would like further information you are welcome to contact either myself or my supervisors:

Zara Godinovich Doctorate of Clinical Psychology student Massey University zara.godinovich@outlook.com 021 1694768	Dr Richard Fletcher Senior Lecturer Massey University r.b.fletcher@massey.ac.nz +64 9 414-0800, Ext 41213	Dr Ian de Terte Senior Lecturer and Clinical Psychologist Massey University i.deterte@massey.ac.nz +64 4 801 5799, Ext 62033

Thank you for reading this information sheet.

Yours sincerely,

Zara Godinovich

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

APPENDIX E-5

CONSENT FORM FOR CHRONOLOGICALLY AGE MATCHED PARTICIPANTS

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

Participant consent form

I have read and understood the participant information sheet for this project. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage. I know that:

1. My participation in the project is entirely voluntary;
2. I am free to withdraw from the project at any time without any disadvantage;
3. Personal identifying information will be destroyed at the conclusion of the project but any raw data on which the results of the project depend will be retained in secure storage for five years, after which they will be destroyed;
4. This project should not cause any discomfort or risks;
5. No formal payment will be offered, however I will receive a voucher of the value of $10.00 (check) as a token of appreciation for my participation;
6. The results of the project may be published and may be made available on the internet but every attempt will be made to preserve my anonymity.

Would you like your GP to be notified if the supervising Clinical Psychologist identifies any potential concerns relating to the autism screening measure. If yes, please list the GP's details below.

☐ Yes ☐ No

Doctor's name: ____________________ Medical centre: ________________________
Contact number: ________________

I agree to participate in this study under the conditions set out in the participant information sheet.

Signature: .... **Date:** ....

Full Name: ....

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

APPENDIX E-6

SCHOOL RECRUITMENT LETTER

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

Research project overview

Thank you for showing an interest in this project. If you decide to participate, we thank you. If you decide not to take part, there will be no disadvantage to you or your school and we thank you for considering our request.

Please note that we have **not** selected your school in the interest of recruiting participants with intellectual disability. Rather children without intellectual disability who fall between the ages of 5 to 12 years have been identified as an appropriate comparison group for the purposes of this research.

Project Summary

People with intellectual disability often experience impairments in social adjustment, as well as the development of psychopathology in addition to their social difficulties. It remains unclear whether reading facial emotions is a main contributor to the social difficulties experienced by this group.

The purpose of this study is to trial a computer based tool designed to test the emotion recognition abilities of people with an intellectual disability. The study will consider how cues such as movement and emotion intensity affect emotion recognition performance, and will also explore the link between emotion recognition and social functioning.

We would like to use this tool to better understand how people with an intellectual disability perform on emotion recognition tasks. To achieve this we need to establish how individuals without intellectual disability would perform on such a task and it is for this reason that we invite your students to participate. We hope that this study will help researchers clarify the link between emotion recognition and social skills, and in time help coordinate better care and learning opportunities for individuals with an intellectual disability.

The coordinating researcher on this project is Zara Godinovich who is completing this study in partial fulfillment of the degree Doctor of Clinical Psychology. The supervisors for this project are Dr Ian de Terte, and Dr Richard Fletcher, both senior lecturers for the School of Psychology, Massey University.

Participants and Recruitment

We are looking for approximately 30 children **without** an intellectual disability to participate in our study. Willing participants will be selected to take part in the study if they have the same receptive language ability as one of the adult participants with intellectual disability already participating in the study. We aim to recruit upward of 300 children in that hope that approximately 30 will be able to be matched to one of our adult participants. Please note that there will be no contact between students from your school and adult participants in the study. Unfortunately we cannot include children with an autism spectrum diagnosis or learning disability.

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

School Involvement

Once I have received your consent to approach students to participate in the study, I will

- Offer a brief talk to students between the ages of 5 and 12 years. Arrangements can be made with individual teachers. I will speak to students about intellectual disability, the purposes of the research and what participation will involve.
- Children who are interested in participating will be issued information sheets and consent forms to take home to their parents.
- I will then arrange a time with your school for data collection to take place.

Project and Procedures

Should you nominate your school to participate the researcher will complete the testing on school grounds. Part B of the study may be completed at the Massey University Centre for Psychology in Albany North Shore City should parents prefer. Consent will be sought from both students and their caregivers.

Part A
The student's receptive vocabulary skills will be assessed. The task requires each student to select one of four pictures that represents a word read aloud by the researcher. This task will be completed in class in a group setting. Students found to have receptive vocabulary skills at a similar level to one of the adult participants in our study, will asked to continue onto Part B of the study. Part B will be completed at a later date and involves one-to-one sessions with each student.

Part B
Parents or caregivers will be asked to complete two measures, an autism screening measure and an assessment of social skills. Each student will complete a series of computerised tasks. The tasks will involve identifying facial emotions, colours, and people's ages. Students may request a break at anytime. Children will be offered a $10.00 NZD voucher (Westfield or Warehouse) as a token of our appreciation. If several classes select to participate it may be possible for a gift to be presented to the school in place of vouchers for individual students. Please let the researcher know if you are interested in this option.

It is important that you are aware that some of procedures outlined above will access receptive vocabulary, and screen for autistic traits. We would like to stress that these are tests used for research purposes and do not offer the same information as a comprehensive clinical assessment. Should individual students score outside the 'normal range' this does not necessary equate to any abnormality or problem. For this reason we do not intend to share these scores with anyone outside of the research team. However, parents are welcome to request that the child's GP be notified should the supervising Clinical Psychologist identify any potential concerns.

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

What is the benefit to my students and our school?

It is our hope that in addition to helping improve our scientific understanding of social functioning in intellectual disability, students will also personally benefit from an increased awareness, sensitivity and acceptance regarding the nature of intellectual disability and its associated challenges. A letter or poster detailing the results of the study and the schools involvement may also be requested. Finally, students are most welcome to ask the researcher questions about her experiences in the field of psychology.

Participant Rights

Your school is under no obligation to accept this invitation. If students from your school decide to participate, they have the right to:

- decline to answer any particular question;
- withdraw from the study anytime up to two weeks post the testing session;
- ask any questions about the study at any time during participation;
- provide information on the understanding that their name will not be used unless permission is given to the researcher;
- request a summary of the project findings when it is concluded.

Data Management

The results of the project may be published and will be made available in the lead researcher's doctoral thesis. Confidentiality will be maintained. We will not share student's personal scores with anyone outside the research team unless they are in a deidentified form.

Ethical Approval

This project has been reviewed and approved by the Northern A Health and Disability Ethics Committee (ref 14/NTA/64).

Project Contacts

If you would like to nominate your school to take part in the study please contact Zara Godinovich. If you have any questions or would like further information you are welcome to contact either myself or my supervisors:

Zara Godinovich
Doctorate of Clinical Psychology student
Massey University
zara.godinovich@outlook.com
021 1694768

Dr Ian de Terte
Senior Lecturer and Clinical Psychologist
Massey University
i.deterte@massey.ac.nz
+64 4 801 5799, Ext 62033

Dr Richard Fletcher
Senior Lecturer
Massey University
r.b.fletcher@massey.ac.nz
+64 9 414-0800, Ext 41213

Thank you for reading this information sheet.

Yours sincerely,

Zara Godinovich

APPENDIX E-7

SCREENING PHASE PARENT INFORMATION SHEET AND CONSENT FORM

FOR MENTAL AGE MATCHED PARTICIPANTS

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

Information Sheet

This letter is to inform you [redacted] School will be taking part in a Massey University research study. We are looking for students to serve as a comparison group in a study focused on social skills and emotion recognition abilities in people with intellectual disability. It is possible that your child's class will be screened to see if any of the students may be suitable to take part. Please note that your child has **not** been selected because they have an intellectual disability.

In order to understand how individuals with an intellectual disability perform on tasks of emotion recognition and social skills, we first need to know how people without intellectual disability perform on such tasks. It is for this reason that we invite students to participant. We need children to take part in our study for the reason that their language abilities are comparable to many adults with intellectual disability. For example, an eight year old child may have the same vocabulary as a 35 year old adult with intellectual disability, and is therefore an ideal participant with whom to compare performance.

Please return the attached form if you would like to withdraw your child from the screening process. If you decide not to take part, there will be no disadvantage to you or your child and we thank you for considering our request.

Project summary

People with intellectual disability often experience impairments in social adjustment, as well as associated psychological difficulties. It remains unclear whether reading facial emotions is a main contributor to the social difficulties experienced by this group. The purpose of the study for which we are screening students, is to trial a computer based tool designed to test the emotion recognition abilities of people with an intellectual disability. We hope that this study will help researchers clarify the link between emotion recognition and social skills, and in time help coordinate better care and learning opportunities for individuals with an intellectual disability.

The coordinating researcher on this project is Zara Godinovich who is completing this study in partial fulfilment of the degree Doctor of Clinical Psychology. The supervisors for this project are Dr Richard Fletcher and Dr Ian de Terte, both senior lecturers for the School of Psychology, Massey University.

What will my child be asked to do?

The screening process will involve an assessment of your child's vocabulary skills using a picture based task. This task will likely be completed in class in a group setting. If your child's vocabulary is found to be similar to one of the adult participants in our study, we will provide you with further information and invite your child to take part in the study. There is no obligation to accept this request. Continued involvement will involve your child completing a series of computer based tasks during class time. We expect this to be a fun and interactive experiment. Please note that students will not have contact with the adult participants in this study.

Participant rights

If you allow your child to participate, both you and your child have the right to: withdraw from the study

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES
AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

anytime; ask any questions about the study at any time during participation; provide information on the understanding that yours and your child's name will not be used unless you give permission to the researcher.

Data management We will not share your child's test scores with anyone outside the research team unless they are in a deidentified form. Data from the screening phase of the study will only be used should your child continue on with the second stage of the study.

Ethical Approval
This project has been reviewed and approved by the Northern A Health and Disability Ethics Committee (ref 14/NTA/64).

Project Contacts
If you have any questions, would like to meet with the researcher, or would like further information you are welcome to contact either myself or my supervisors:

Zara Godinovich Doctorate of Clinical Psychology student Massey University zara.godinovich@outlook.com 021 1694768	Dr Ian de Terte Senior Lecturer and Clinical Psychologist Massey University i.deterte@massey.ac.nz +64 4 801 5799, Ext 62033	Dr Richard Fletcher Senior Lecturer Massey University r.b.fletcher@massey.ac.nz +64 9 414-0800, Ext 41213

Thank you for reading this information sheet.

Yours sincerely,

Zara Godinovich

NON-CONSENT FORM

Please **do not** include ____________________________*(student's name)* from Room ___ , in the screening phase of the "Assessing Facial Affect Recognition in People with Intellectual Disability" research study.

Full Name: ___________________________________ **Signature:** _______________

Relationship with student: _______________________ **Date:** ____/____/ 2014

Contact number: _______________________

APPENDIX E-8

EXPERIMENTAL PHASE PARENT INFORMATION SHEET FOR MENTAL AGE MATCHED PARTICIPANTS

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

Invitation

[redacted] School is currently participating in a Massey University research study. Students have recently been screened on their verbal abilities to see if they are a fitting candidate for this research. Based on their scores on a vocabulary test, your son or daughter has been found to be the ideal candidate to serve in a comparison group, in a study focused on social skills and emotion recognition abilities in adults with intellectual disability. Please note that your child has been asked to take part because they **do not** have an intellectual disability.

Project Summary

People with intellectual disability often experience impairments in social adjustment, as well as associated psychological difficulties. It remains unclear whether reading facial emotions is a main contributor to the difficulties experienced by this group. The purpose of the study is to trial a computer based programme designed to test the emotion recognition abilities of people with an intellectual disability. We hope that this study will help researchers clarify the link between emotion recognition and social skills, and in time help coordinate better care and learning opportunities for individuals with an intellectual disability.

In order to understand how individuals with an intellectual disability perform on tasks of emotion recognition and social skills, we first need to establish how people without intellectual disability perform on such tasks. It is for this reason that we invite students to participant. We need children to take part in our study for the reason that their early language abilities are comparable to many adults with intellectual disability. Your child has been found to have similar verbal abilities to one of the adults in our study. Please note that due to the range of abilities apparent in individuals with intellectual disability, children of varying abilities have been selected for this study. Your child will not have contact with any of the adult participants in this study.

Please return the attached form to your child's teacher, indicating whether you and your child would like to be involved in this study. If you decide not to take part, there will be no disadvantage to you or your child and we thank you for considering our request. Please do return the form even if you do not wish to participate, so that we may invite another student to take part. For those who do not wish to involve their child in this study, we are also looking for adult participants without intellectual disability to take part. For further information about this please contact the coordinating researcher.

The coordinating researcher on this project is Zara Godinovich who is completing this study in partial fulfillment of the degree Doctor of Clinical Psychology. The supervisors for this project are Dr Richard Fletcher and Dr Ian de Terte, both senior lecturers for the School of Psychology, Massey University.

What will we be asked to do?

Your child will complete a series of tasks during school time. The tasks will involve identifying facial emotions, colours, and people's ages using an iPad. Students will be offered a $10.00 NZD

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

Warehouse voucher as a token of our appreciation. We hope to make all testing sessions light spirited, fun, and interesting for all students involved.

You (the parent or caregiver) will be asked to complete two questionnaires about your child. This information is completely confidential and may be completed at your leisure. The first questionnaire will focus on your child's social skills and the second is an autism screening measure. We would like to stress that these questionnaires are used for research purposes only, and do not offer the same information as a comprehensive clinical assessment. Should your child score outside the 'normal range' this does not necessary indicate any abnormality or problem. For this reason we do not intend to share this information with anyone outside the research team. However, you are welcome to request that the child's GP be notified should the supervising Clinical Psychologist identify any potential concerns.

Participant rights

If you allow your child to participate, both you and your child have the right to:
- decline to answer any particular question
- provide information on the understanding that yours and your child's name will not be used unless you give permission to the researcher
- withdraw from the study anytime up to two weeks after the testing session
- request a summary of the project findings when it is concluded

Data management

The results of the project may be published and will be made available in the lead researcher's doctoral thesis. Confidentiality will be maintained. We will not share your child's personal information or scores with anyone outside the research team unless they are in a deidentified form.

Ethical Approval

This project has been reviewed and approved by the Northern A Health and Disability Ethics Committee (ref 14/NTA/64).

Project Contacts

If you have any questions, would like to meet with the researcher, or would like further information you are welcome to contact either myself or my supervisors:

Zara Godinovich
Doctorate of Clinical Psychology student
Massey University
zara.godinovich@outlook.com
021 1694768 (text/call 10am-4pm)

Dr Richard Fletcher
Senior Lecturer
Massey University
r.b.fletcher@massey.ac.nz
+64 9 414-0800, Ext 41213

Dr Ian de Terte
Senior Lecturer and Clinical Psychologist
Massey University
i.deterte@massey.ac.nz
+64 4 801 5799, Ext 62033

Thank you for reading this information sheet.

Yours sincerely,

Zara Godinovich

APPENDIX E-9

EXPERIMENTAL PHASE PARENT CONSENT FORM FOR MENTAL AGE MATCHED PARTICIPANTS

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

Parent/caregiver consent form

I have read and understood the Letter of Invitation concerning this project. All my questions have been answered to my satisfaction. I understand that I am free to request further information at any stage. I know that:

1. Participation by my child in the project is entirely voluntary;
2. I will need to complete a series of questionnaires about how my child interacts with others;
3. Only children who consent and whose legal guardians consent will participate in the project;
4. Students may withdraw from the study at any time without any disadvantage;
5. Personal identifying information will be destroyed at the conclusion of the project but any raw data on which the results of the project depend will be retained in secure storage for five years, after which they will be destroyed;
6. This project should not cause any discomfort or risks;
7. No formal payment will be offered, however my child will receive a voucher for the value of $10.00 as a token of appreciation for their participation;
8. The results of the project may be published and may be made available on the internet but participant anonymity with be preserved.

Would you like your child/dependent's GP to be notified if the supervising Clinical Psychologist identifies any potential concerns relating to the autism screening measure? If yes, please list the GP's details below.

☐ Yes ☐ No

Doctor's name: ____________________ Medical center: ________________________
Contact number: ____________________

I **DO / DO NOT** (please circle) give consent for you to approach ____________________ *(insert child's legal name)* from Room __ to participate in the project "Assessing Facial Affect Recognition in People with Intellectual Disability".

Full Name: ________________________________ Signature ______________

Relationship with student: ______________________ Phone: ______________

Postal Address (to send questionnaires): ______________________ Date: ____/____/ 2014

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

APPENDIX E-10

EXPERIMENTAL PHASE PARTICIPANT INFORMATION AND CONSENT FORM FOR MENTAL AGE MATCHED PARTICIPANTS

MASSEY UNIVERSITY
COLLEGE OF HUMANITIES AND SOCIAL SCIENCES
TE KURA PŪKENGA TANGATA

Assessing Facial Affect Recognition in People with Intellectual Disability

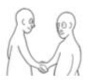
Participant information summary and consent form

This study is to learn more about how people with an intellectual disability recognise emotions on other people's faces. To learn about people with intellectual disability it is helpful for us to compare them with people without an intellectual disability. This is why I have asked you to take part in the study. **You do not have an intellectual disability.**

You will start by doing an activity with the rest of your class where the researcher will show you four pictures and then read out a word. You will need to point to the picture that matches the word. You may or may not then be asked to do some activities on the computer with the researcher. This will probably happen on a different day.

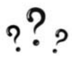
You will be shown pictures and videos of people's faces on a computer. You will be asked to guess how they are feeling or to guess how old they are. You will also be shown some moving shapes and will be asked what colour they are.

We will write about what we find out from working with you. This may be published in a scientific journal, which is like a book. We will not put your name in the journal so no one will know who you are.



We will ask your parents or caregiver some questions about how you interact with other people.

If you get tired you can ask for a break.



If you have any questions you can ask me at anytime.

If you need to complain about anything, please tell me or talk to your teacher.

You will get a $10 voucher for taking part.

It is your decision to take part — You can say "Yes" 

You can say "No" 

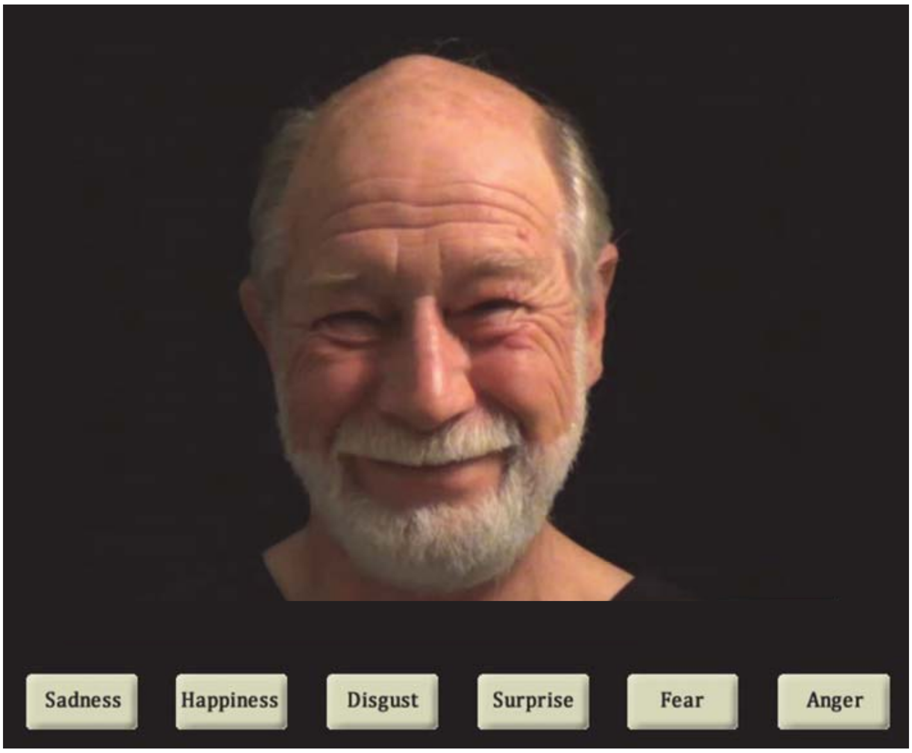
If you understand what I have told you and you want to take part in the study please sign your name.

Participant ____ Date __/__/__

Massey University, School of Psychology – Te Kura Hinengaro Tangata
PO Box 756, Wellington 6140, New Zealand
W: www.massey.ac.nz T: +64-4-8015799

APPENDIX F

EXPERIMENTAL TASK INTERFACE



APPENDIX G

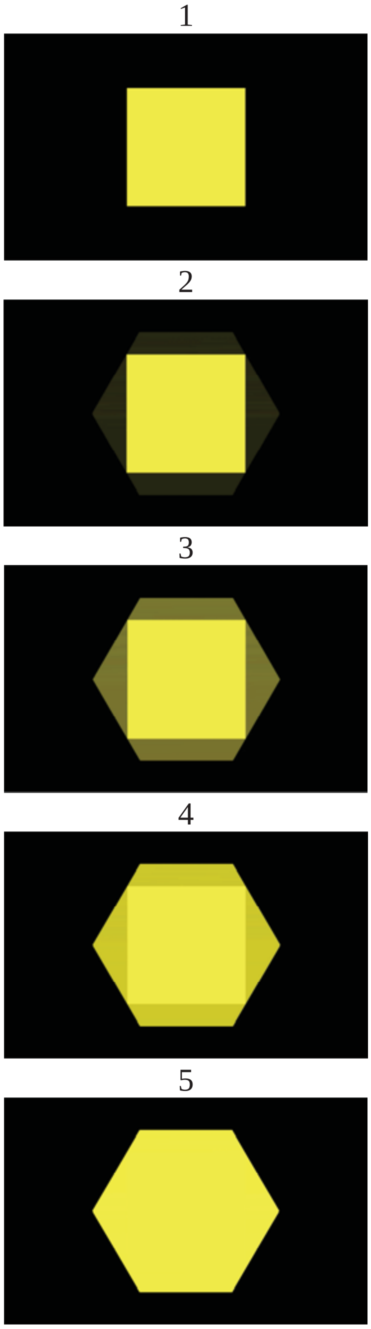
SELECTED POFA, NIMSTIM AND DAFEX STIMULI

	PoFA	NimStim [a]	DaFEx [b]
Anger	MF2-7	Actor 31, C	Actor 2, B3
	JB1-23	Actor 2, O	Actor 4, B3
	JM5-3	Actor 42, C	Actor 7, B3
Disgust	PF1-24	Actor 9, C	Actor 5, B6
	WF4-22	Actor, 27 C	Actor 6, B3
	CI-4	Actor 41, O	Actor 8, B6
Fear	GSI-25	Actor 10, O	Actor 3, B6
	PE3-21	Actor 5, O	Actor 2, B6
	MO1-26	Actor 36, O	Actor 7, B6
Happiness	WF2-12	Actor 14, O	Actor 1, B3
	EM4-7	Actor 6, O	Actor 5, B3
	A1-6	Actor 21, O	Actor 6, B3
Sadness	PE5-7	Actor, 1, C	Actor 1, B6
	NR2-15	Actor 7, C	Actor 6, B6
	JJ5-5	Actor 40, C	Actor 8, B3
Surprised	A1-24	Actor 17, O	Actor 3, B3
	GSI-16	Actor 23, O	Actor 4, B3
	SW1-16	Actor 35, O	Actor 5, B3

a. *O* = Open mouth, *C* = Closed mouth
b. *B* = Block

APPENDIX H

EXEMPLAR OF THE COLOUR DISCRIMINATION TASK ANIMATION AT FIVE TIME-POINTS



APPENDIX I

EFFECT SIZE INTERPRETATION THRESHOLDS

Test	Relevant effect size	Effect size threshold		
		Small	Medium	Large
Cohen's *d*	d	.20	.50	.80
Omega squared	ω^2	.01	.06	.14
Kendall's Coefficient of Concordance	W	≤.20 (very weak) .20-.40 (weak)	.40-.60	strong .60-.80 (strong) .80- 1.00 (very strong)
Pearson product-moment correlation and r = Z/√N	r	.10	.30	.50

Note. Proposed thresholds for Pearson product moment correlation and Cohen's d from Cohen (1988), for r = Z/√N from Pallant (2007), for omega squared from Field (2005), and Kendall's Coefficient of Concordance from Rovai, Baker, & Ponton (2013).

APPENDIX J

GROUP EMOTION RECOGNITION TASK PERFORMANCE ORGANISED BY TASK AND EMOTION CATEGORY

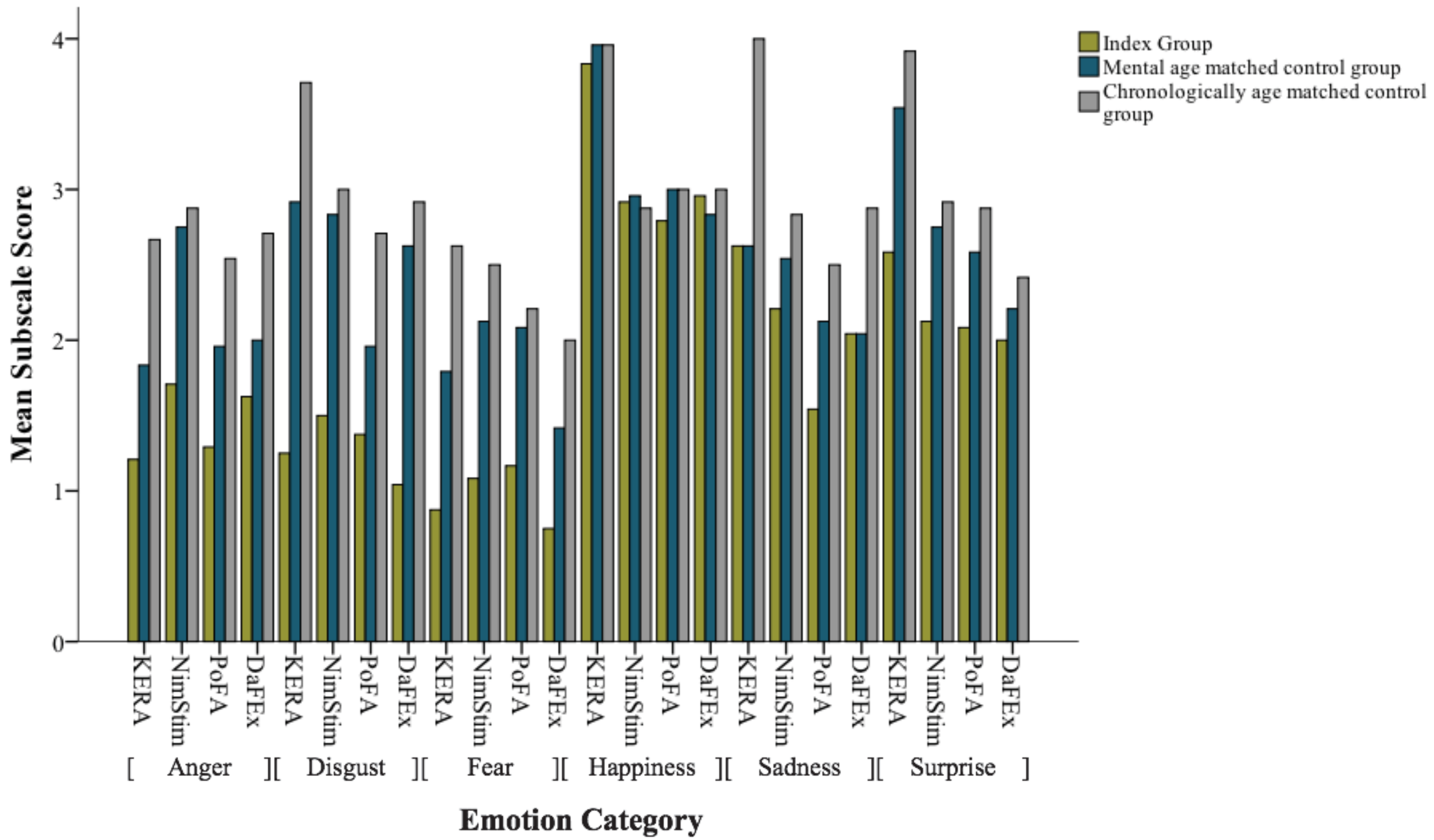


Figure J.1. Group emotion recognition task performance scores organised by stimuli type and emotion category