



# The New Zealand Journal of Psychology

Volume 51, Issue 2 (August), 2022  
ISSN:1179-7924



**New Zealand  
Psychological Society**  
*Rōpū Mātai Hinengaro o Aotearoa*

---

# New Zealand Journal of Psychology

---

## EDITOR

**Marc Wilson**

*School of Psychology  
Victoria University of Wellington*

## ASSOCIATE EDITORS

**Neville Blampied**

*Department of Psychology  
University of Canterbury*

**Janet Leathem**

*School of Psychology  
Massey University at Wellington*

**Harlene Hayne**

*Department of Psychology  
University of Otago*

**Michael O'Driscoll**

*Department of Psychology  
University of Waikato*

## BOOK REVIEW EDITOR

**Iris Fontanilla**

*Auckland DHB*

© The material published in this issue is  
copyright to the New Zealand Psychological  
Society.

*Publication does not necessarily reflect the  
views of the Society*

## The New Zealand Journal of

**Psychology** is published online in three  
editions a year (articles will be posted as  
they become available) by the **New  
Zealand Psychological Society Inc.**

*Content may include manuscripts and  
shorter research notes in any substantive  
area of Psychology, and book reviews.  
Submitted manuscripts may be (1)  
empirical, (2) reviews of the literature, or  
(3) discussions of theoretical/conceptual  
frameworks of relevance to the practice of  
Psychology. Manuscripts will be  
considered for publication if they (a)  
include data collected from New Zealand  
samples, or (b) discuss the relevance of  
issues contained in the manuscript to the  
New Zealand social and cultural context,  
or to the practice of Psychology in this  
country. For further clarification of these  
requirements, please contact the Editor.*

See **Instructions to Authors** at:

<https://www.psychology.org.nz/publication-s-media/new-zealand-journal-of-psychology/#author-guidelines> for  
preparation and submission of  
manuscripts and material.

## Subscriptions

Subscription to the Journal is included in  
Membership dues for the Society. Non-  
members may subscribe at NZ\$55.00 per  
volume/year (three issues), plus GST (NZ  
only). Subscription and advertising  
enquiries should be addressed to the:

*Business Manager  
New Zealand Journal of Psychology  
NZ Psychological Society Inc.  
(contact details below)*

## **Production, Printing & Distribution**

Is managed by National Office

*NZ Psychological Society Inc.  
P.O. Box 10536, The Terrace,  
Wellington 6143,  
New Zealand  
Ph (04) 473 4884  
Email: office@psychology.org.nz*

---

---

## Editor's Introduction

---

**An introduction to this issue, a reminder of the upcoming special issue on environment and climate, and some advice for authors.**

**Prelude to this issue:** Volume 51, issue 2 presents six articles, including authors from five of Aotearoa's seven Universities, and representing a breadth of research and practice. The topics are diverse, from neuropsychological norms, through social-political psychology, to statistical smackdowns between different theoretical frameworks. This work includes qualitative investigations (Gallagher and colleagues interviews of ten men about understandings of mental health services; Turner-Adams and Webber's interviews and surveys of a large number of mentoring programme stakeholders), quantitative analyses (Dudley and colleagues presentation of norms from an impressive sample of 284 Māori; Kappmeier and Fahey's analyses of surveys About trust, and; Jackson and colleagues contrasting of application of G Theory against confirmatory factor analysis) as well as Amersfoort and Friesen's mixed-methods investigation of surveys with a large group of users of a postnatal support programme. The authors affiliations locate them not just in Schools of psychology, but also Health Sciences, Education (and Education and Social Work), Peace and Conflict Studies, Māori Health, and Biostatistics and Epidemiology. This is a methodological and sub-disciplinary diversity that I'd like to continue to encourage. If you have work that you'd like to have considered by the Journal, but you're unsure of suitability, please contact us.

I shall comment on two of these works in particular. First, while the data that Kappmeier and Fahey analyse is drawn from sample of Bostonians, it illustrates something essential for NZJP – establishing relevance to our own context. The majority of desk rejections handled by NZJP are rejected for failing to make this case. Secondly, while Jackson and colleagues aren't located in New Zealand at all (though the team includes New Zealanders), they present a comparison of different theoretical frameworks illustrating an enviable New Zealand-based sample of people located in organisations. It's worth noting that a quick survey of the most highly cited papers ever published in NZJP include an over-representation of research coming from an industrial-organisational perspective.

**Last chance - upcoming special issue on psychological perspectives on environment, climate and sustainability:** We are close to completing review and acceptance of some of the manuscripts submitted for this special issue and will publish individual papers as they become available, as well as in combination in a supplementary December issue of the Journal. While it's extremely late in the piece, we shall continue to operate an expedited review process for work that is potentially relevant to this theme.

We strongly encourage scholars with work relevant to the subject of this special issue to consider submission. For further detail please contact Marc Wilson. Special issue Editors are drawn from the New Zealand Psychological Society's Climate Psychology Task Force and include Brian Dixon, Jackie Feather, Natasha Tassell-Matamua, and Marc Wilson. For further information about the Society's Climate Change initiatives please visit the Society [website](#).

**Advice to authors:** Consistent with the imperative of the Journal, *any* submission must clearly articulate relevance in the context of Aotearoa New Zealand. Information about the Journal, and general author guidelines can be found [here](#).

Additionally, the Covid-19 pandemic has dramatically affected the ways that academics work, and this can be seen in much greater difficulty securing reviews (for example, we have experienced significantly more declines of review invitations compared to pre-Covid times). Feel free to suggest reviewers with appropriate expertise (while being aware of conflicts of interest) and we will draw off that list when supplementing the invitations we extend. Finally, **please ensure that you submit a deidentified manuscript!**

**Marc Wilson**

---

---

# New Zealand Journal of Psychology

Volume 51, Number 2 (August), 2022

## CONTENTS

"Do They Chain Their Hands Up?": An Exploration of Young Men's Beliefs about Mental Health Services	<i>Jake M. Gallagher, Keith Tuffin, and Clifford van Ommen</i>	4
Age stratified normative data for Māori on the Wechsler Adult Intelligence Scale (4th edition; WAIS-IV)	<i>Makarena Dudley, Suzanne Barker-Collo, Denise Wilson, and Nick Garrett</i>	15
Stakeholders' Perceptions of a New Zealand Youth Mentoring Programme Assisting High-Achieving, Underprivileged Students to Attend University	<i>Hana Turner-Adams and Melinda Webber</i>	26
Trust and legitimacy: Policing among racial groups	<i>Mariska Kappmeier and Kathryn H. Fahey</i>	35
Space for you and your baby: Participant perceptions of community-based postnatal parenting support and adjustment to parenthood	<i>Linde-Marie Amersfoort and Myron Dean Friesen</i>	43
Clarifying the Scope of Generalizability Theory for Multifaceted Assessment	<i>Duncan J. R. Jackson, George Michaelides, Chris Dewberry, and Paul Englert</i>	53

## “Do They Chain Their Hands Up?": An Exploration of Young Men's Beliefs about Mental Health Services

Jake M. Gallagher<sup>1</sup>, Keith Tuffin<sup>1</sup>, and Clifford van Ommen<sup>2</sup>

<sup>1</sup>School of Psychology, Massey University, Wellington

<sup>2</sup>School of Psychology, Massey University, Auckland

Young men tend to be less likely to seek help for mental distress due to barriers including beliefs about mental health services. However, little research examines beliefs of men who have not accessed services. In the present study, ten young men who had not utilised services participated in interviews regarding their beliefs about mental health treatment. Data were analysed with inductive thematic analysis within a social constructionist epistemology, resulting in five themes. Overall, participants expressed some trepidation about utilising services, and were aware that their knowledge was limited. Participants expressed a preference to fix problems independently, negative views about relying on prescription medication, and they likened talk-therapy to informal social support. Participants also acknowledged the limit of their beliefs, which were based on fictional depictions. It was concluded that better public education regarding treatment may reduce barriers to help-seeking.

**Keywords:** *Help-seeking, mental health services, young men, masculinity, New Zealand*

### INTRODUCTION

Men typically demonstrate low rates of mental health service use (Pattyn et al., 2015; World Health Organization, 2002). This being the case despite higher rates of completed suicide for men compared to women indicating a significant need amongst men for such services (New Zealand Ministry of Health, 2016). Additionally, Clement et al. (2015) and Keown et al. (2016) found that young men in lower socio-economic groups experience high distress yet are among the least likely to seek help. Likewise, Lynch and colleagues (2018) noted that young men have particularly high suicide rates and low rates of help-seeking (relative to other groups) and found that young men experience many barriers to help-seeking.

Research has typically focussed on the notion of barriers as a key concept in understanding lower rates of health service utilisation amongst men. Barriers to help seeking may be classified as ‘attitudinal,’ referring to attitudes and beliefs, or ‘structural,’ relating to practical factors such as cost, distance, or time constraints (Andrade et al., 2014). Structural barriers have, unsurprisingly, been found to have a greater impact on rates of help-seeking among those unable to afford services, whereas attitudinal barriers affect men, irrespective of the degree to which they are impacted by structural barriers (Rice et al., 2018; Walker et al., 2015). The aim of the present study is on better understanding these attitudinal barriers.

Attitudes and beliefs are important concepts used by researchers in understanding mental health service utilisation amongst men (Yousaf et al., 2015). Understood in various ways, Fazio (1986) defines attitude as the “categorisation of an object along an evaluative dimension” (p. 214), with Petty (2018) further distinguishing attitudes, as comprising the emotional valence towards an object, from beliefs, as comprising

statements – whether true or not – about the object. Attitudes and beliefs are seen to influence one another with substantial evidence linking attitudes to behaviours and vice versa – though this relationship is not perfectly predictive (Maio et al., 2018). Accordingly, men who hold negative attitudes towards mental health services, and who believe that services may be harmful or unhelpful, would be less likely to seek help.

Reviewing the literature, Yousaf et al. (2015) found that men's understanding of services, fear of diagnoses, and previous unhelpful experiences with service providers were barriers to help-seeking. Generally, research indicates that the public have a mixed understanding of talk therapy (Cramer, 1999), which would impact men's utilisation of services (McKelley & Rochlen, 2007; Rice et al., 2018). Specifically, Coles et al. (2010) found that some men expressed a general distrust of health services and were frustrated by the lack of appropriate services, whilst other studies indicate concerns about confidentiality (Gonzalez et al., 2005). Some men believed that pharmaceutical treatment was the only intervention available (House et al., 2018), with some believing that these drugs were dangerous and addictive (Lauber et al., 2005; Mirnezami et al., 2016).

Schultz (2005) argued that due to low rates of service use, public beliefs and attitudes towards specialist mental health services are often informed by popular media depictions, which tend to use extreme characterisations (Orchowski et al., 2006; Wedding 2017). Vogel, Gentile, and Kaplan (2008) found that people who reported watching more content relating to mental health services expressed greater fear and reduced confidence towards services. In contrast, men who *had* utilised services tended to endorse more positive views (Harris et al., 2016; Sierra et al., 2014). Considering New Zealand (NZ), men's mental health beliefs are also likely influenced by local

media with news reporting and campaigns impacting in ways not accounted for in overseas research. For example, Sir John Kirwan's campaign of normalising and educating men about depression has been influential (Wardell, 2013). NZ men's beliefs and attitudes would also be influenced by local cultural, family, and religious beliefs regarding mental health services (Lynch et al., 2018; Vogel et al., 2007).

Harding and Fox (2015) aimed to understand what men who had sought help believed were the key factors that enabled them to do so. Based on interviews with nine men, they found that prior to help-seeking, the men had been worried about services comprising "Freudian couches and personality changing drugs" (p. 457). Harding and Fox reported that all these men had previously had negative understandings of treatment and had felt relieved when actual treatment did not match their negative expectations.

Analysing interviews and focus groups of young men utilising a service in the North West of Ireland, Lynch et al. (2018) articulated several barriers to help seeking, including perceived and actual negative peer, community, and medical profession reactions, difficulty in articulating emotions and problems, a sense of compromised self-reliance and masculinity, the use of ineffective coping mechanisms (such as alcohol consumption), conservative religious norms, and an expectation of unsympathetic incomprehension from older generations.

Although the above studies provide useful insights, these participants' views were informed by the experience of service use. In contrast and focussing specifically on talk therapy and no other aspects of services, Midgley et al. (2016) interviewed young people about their beliefs of what this would entail and found that most participants did not know what might happen.

The above attitudes and beliefs can be usefully placed within a broader conceptual framework of gender and, more specifically, hegemonic masculinity. Separating the biological facticity of sex from the sociocultural embeddedness of gender, researchers have foregrounded the latter as a predominant factor in considering help seeking amongst, especially young, men (e.g., Cleary, 2012; Moller-Leimkuhler, 2002; Vogel et al., 2011). Research has indicated that men experience pressures to behave according to hegemonic masculine norms regardless of ethnic culture or other cultural categories (Ramaeker & Petrie, 2019; Tan et al., 2013).

Hegemonic masculinity refers to enduring gender-related power dynamics and behaviours typically privileging stereotypical masculine characteristics. These may include stoicism, an emphasis on self-control, and dominating those who exhibit stereotypical feminine traits such as openness and emotionality (Jewkes et al., 2015). Hegemonic masculinity; however, also needs to consider cultural variability. For example, Hamley and Le Grice (2021) claim that, prior to colonisation and allowing for cultural heterogeneity, Māori gender roles allowed for men in various roles to flexibly adopt both masculine and feminine behaviours. This contrasts with the rigidity associated with traditional western gender roles.

Hegemonic masculinity can thus be understood as preventing or delaying help-seeking, particularly for mental health concerns, as such behaviour is regarded as

an incompatible feminine behaviour (Krumm et al., 2017). Seidler et al. (2016), for example, found that men who attempt to behave according to the dictates of hegemonic masculinity were less likely to seek help for depression. Although research has linked poorer mental health literacy and understanding of symptoms to delayed help-seeking irrespective of gender (Jorm, 2000, 2012), there is also evidence of a gender effect. Levant et al. (2009) found that alexithymia (i.e., an inability to articulate feelings) was associated with men who identified with hegemonic masculine values, suggesting that such men would have difficulty in articulating symptoms of mental distress. Accordingly, Swami (2012) found that men tend to have poorer understanding of common disorders such as depression. Interestingly, research indicates that men with poor mental health literacy also have a poor understanding of, and negative attitudes towards, associated services (Jorm, 2012), suggesting that even if symptoms are recognised, men may *choose* not to seek help. Thus, difficulty in recognising and articulating symptoms, and fear of stigma for experiencing mental distress or utilising services, are identified as barriers to help-seeking among men (Cleary, 2012; Clement et al., 2015).

The present study aims to address the question of what young men who have not accessed services believe such services entail and what factors act as barriers to such utilisation. Despite the relevance of attitudes and beliefs as a barrier to help-seeking being well established, further understanding is needed of what comprise men's attitudes and beliefs. Furthermore, since much of the extant literature have been quantitative studies (e.g., Coles et al., 2010; Furnham, 2009; Harris et al., 2016), further qualitative research is warranted in providing a more nuanced understanding of men's experiences and meaning making regarding services.

## METHOD

### Design

This study aimed to explore, using open-ended interview questions, men's beliefs and attitudes towards mental health services. An exploratory inductive, qualitative interview design was used.

The concepts formulated in this research were considered through a social constructionist epistemology. That is, data in this study were considered as products of time, place, and circumstance (Burr, 2015; Gergen, 1985). For example, the concept of 'attitudes' and 'beliefs' are viewed as being constructed in conversation and serving a social-contextual function. Furthermore, as Tuffin and Danks (1999) argue, the terms 'attitude' and 'belief' are used as pragmatic codes signalling the discussion of services and gender related issues, rather than as reified phenomena.

A 'reflexive' Thematic Analysis (TA) design based on Braun and colleagues' (2019) description was undertaken in this study as it theoretically aligns with the constructionist epistemology. Such a style acknowledges and embraces the author's own social context; at the time of collecting and analysing data, I was a 28-year-old urban based New Zealand European male, intern psychologist, and someone who had recovered from mental distress without accessing services. This reflexive design accepts

that inherent researcher subjectivity adds richness and context to the data (Braun et al., 2019).

Men's beliefs and attitudes regarding services were constructed in the context of semi-structured one-to-one interviews. Interviews were chosen to facilitate a flexible, iterative process of data generation.

### Recruitment

This study was advertised via a custom-made Facebook page in January 2019. An advertisement was posted on a local public Facebook group with many members. The advertisement included the offering of a supermarket voucher thanks for the sharing of time and knowledge. Interested candidates contacted the researcher directly and were provided with the information sheet and consent forms.

Eligible participants included men between the ages of 18 and 30, who had not utilised mental health services and who were not working in or studying a mental health related field, as previous research has focussed on men who have accessed services (e.g. Harding & Fox, 2015). Whether participants had previously utilised services was based on their own definition of mental health services. It was recognised that young men may not have a clear definition of mental health services and therefore allowing participants to decide whether they had accessed services facilitated conversations regarding definitions of services and provided useful qualitative information. The age criteria were selected due to low rates of service use and growing suicide rates amongst young men (Keown et al., 2016; Lynch et al., 2018). Likewise, young men were selected, as previous research has indicated that young men are less likely to seek help than women, and more likely to minimise health issues than older men (Kessler et al., 1999; Moller-Leimkuhler 2002). There were no exclusion criteria relating to other demographic factors.

**Table 1.** Demographic descriptions of participants

<b>Arjun</b> was a 29-year-old straight Indian male, who was working as a head chef.
<b>Timothy</b> was a 21-year-old straight 'white' male, who was unemployed at the time of his interview.
<b>Tane</b> was a 20-year-old male of Maori, Indian, and European descent. Tane was planning to study to become a personal trainer at the time of interview.
<b>Kris</b> was a 19-year-old bi-sexual New Zealand European male, who was a student and a seasonal nature guide.
<b>Simon</b> was a 24-year-old gay New Zealand European male, who worked as a night auditor in a hotel.
<b>Matt</b> was a 28-year-old African male, who worked in hospitality.
<b>Manish</b> was a 28-year-old straight Indian male, who was a post-graduate university student, and previously worked in a bank.
<b>Geoff</b> was a 22-year-old straight European New Zealand male, who was a student and tutor at university.
<b>Wiremu</b> was a 26-year-old gay Maori European male, who worked as a regulation advisor.
<b>Phil</b> was a 26-year-old gay Samoan male, who worked as a policy analyst.
<b>Wiremu</b> is a 26-year-old gay Maori European male, who worked as a regulation advisor.
<b>Phil</b> is a 26-year-old gay Samoan male, who worked as a policy analyst.

### Participants

Ten men who met inclusion criteria were interviewed. Although the focus was on gender, other aspects of participants' demographic details are recognised here. All participants identified as cisgender men and lived in the Wellington region of New Zealand. Brief demographic descriptions of all participants, using their own words and pseudonym names, are provided in Table 1.

### Data collection

Interviews followed a semi-structured schedule relating to understandings of mental health services; however, participants were encouraged to discuss topics of importance to them. Participants were asked what they believe mental health care in New Zealand involves, eliciting their understandings of types of services, and conversations around these services followed. Prompts included questions such as 'how effective is this form of treatment?' or 'how effective is this for men?' Participants were also asked whether anyone they know has ever experienced significant mental distress, and how they coped. If conversations arose about participants' own intentions or attitudes towards help-seeking and specific services, these topics were explored. Interviews ranged in duration between 26 and 53 minutes, with an average length of 39 minutes. Interviews were conducted in February 2019.

Interviews were conducted at a semi-private location of the interviewee's choice and agreed upon by the research team; nine interviews were conducted in private meeting rooms at university libraries, and one interview was conducted in a private meeting room in a government building. Interviews were digitally recorded and transcribed by the lead researcher/interviewer - a male, and trainee clinician at the time of data collection. Transcription followed an orthographic style, whereby the standard spelling of words was used. Data was not modified for grammatical sense or length to preserve the organic 'feel' of the discussion. This method of transcription was based on the recommendations of Braun and Clarke (2006; 2012), who noted that orthographic transcription that reflects the content of speech in a generally realistic and readable manner is adequate for TA.

### Data Analysis

Reflexive Thematic Analysis (TA) was selected as the method for data analysis, due to its value in structured synthesis of data and generation of themes in relation to an open research question (Braun & Clarke, 2012). Data were analysed following Braun et al.'s (2019) six phases of thematic analysis: becoming familiar with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. This allowed for a clear, consistent, replicable methodological approach. In Reflexive TA, the primary researcher(s) inductively codes data. This method is unlike other forms of TA such as 'coding reliability,' in which inter-rater reliability is sought through enlisting

multiple coders. Braun argued that methods such as coding reliability seek to fit with quantitative methods and are less appropriate than reflexive TA in constructionist methodologies.

Familiarisation with the data commenced during interviews, as initial ideas were generated, and this step continued through into the next steps of data review and transcription, as possible relevant points were noted. The process of generating initial codes flowed through from familiarisation as shorthand codes were noted on transcripts. Initial codes labelled and begin to categorise data in an iterative process. Codes were also collated on a Microsoft Excel spreadsheet. Over all transcripts, 150 initial codes were generated during this process. The process of searching for, and reviewing themes, was also iterative. Initially codes were reviewed for similarities and grouped according to possible overarching themes. Data fitting within these initial themes was then collated and reviewed resulting in further refinement of thematic groupings until a coherent theme evidenced in the data was apparent to the researcher. During this process, several codes were elevated to tentative themes due to richness of data within some codes. Likewise, there were some interesting pieces of data that were not elevated to the level of theme or were omitted from the final report. Themes were then given tentative names, which were finalised during the report writing and as the theme was further defined. The most relevant and useful themes were identified and described in the report writing phase, and data were considered to represent dynamic and contextually bound constructions of meaning about services, rather than objective truths.

### **Ethical Considerations**

This study was assessed through the Massey University research ethics process and deemed 'low-risk.' As a low-risk study, this research was subject to peer-review by a researcher not involved in this study prior to commencement.

Participants were provided with an information sheet and consent form explaining the purpose and scope of the research and freedom to withdraw at any time. Participants' mental health status was not discussed, as this study related to men who were not service users, and it was deemed outside of the scope of this study to assess mental health status. However, participants were also provided with a list of mental health services should they have concerns about their mental health. The interviewer was a trainee clinician under the supervision of a senior clinician, with whom all interviews were discussed. There was no expectation that participants would be at risk, as participants were from the general population, not a clinical population. Additionally, participants were informed that pseudonyms would be used to protect their privacy in the final report or publications and that all data would be securely stored.

To respect Treaty of Waitangi (Treaty) principles regarding engagement of Māori (New Zealand indigenous people) in research, Hudson and Russell's (2009) guidelines regarding Treaty principles were followed. Based on these guidelines, it was deemed essential that the study include participant(s) who identified as Māori. Additionally, this research was ethically assessed by a

Māori researcher, allowing input into the research plan. Furthermore, Māori cultural values were included in the plan to engage with participants. Specifically, participants were offered an option to open and close the session in a way that would be most comfortable for them, and a karakia (blessing/transition) was offered. Additionally, shared kai (food) of biscuits was brought to each interview. Tangible benefits of this research to Māori may include better understanding of barriers to help-seeking in a New Zealand context, and possible improvements made based on the knowledge of these barriers.

### **ANALYSIS AND COMMENTARY**

Analysis of the data resulted in identification of five themes: 1. *'This is all based off what I know from television.'* 2. *'So, it would be some form of social aspect to it.'* 3. *'Maybe just deal with it yourself first.'* 4. *'You don't exactly know how to deal with coming off.'* 5. *'Cause they're the doctor, they know what's best for you.'*

#### ***'This is all based off what I know from television.'***

Participants noted that their knowledge of mental health services was based on movies and television shows and they lacked alternative information sources. In the following extracts, participants discussed what they believe happens in inpatient mental health services:

(1) *Phil: Um do they... uh... I'm just thinking (laughs) like scenes from movies like-*

*J: Yeah, no go ahead, like go ahead if that's what comes to mind -*

*Phil: Do they lock them - you know like - chain their hands up or um do they put them in a room where there's like no knives or something like that - you know? (laughs)*

(4) *Arjun: (...) You know, um I don't - I - I don't think that screaming and shouting and being physical and aggressive and um you know - having some kind of traumatic experience is going to help a person in a situation like that. Um even though that person may be - you know - being wild and like - you have to understand they're not in the correct space uh mentally so... I - I don't think being very aggressive is - is and maybe that's just an assumption cause of movies but (hah) um I hope that's not the case in real life.*

Phil and Arjun described extreme evocative scenarios of inpatient mental health services. They both acknowledged their inexperience by pointing to two dimensional fictional visual accounts, which speaks to the lack of alternative and more nuanced readily available sources of information. Although participants acknowledged the limited validity of media as a source of information, they were forced to draw upon these sources when expressing their impressions of inpatient services. Media informed views were framed as being inaccurate, yet their representations were most available to participants.

As well as inpatient services, participants drew upon images from mass media when describing their image of what happens in talk-therapy:



*Geoff: Um I... Oh... okay, this is all based off what I know from television, basically, so - you know, there'd be a nice big comfy couch... and then the other person would be sitting on a comfier couch, something like that and sort of you would either sit down or lay down or something and you'd probably just start exploring like - you know - how you're feeling... (...)*

Using what they had seen in film and television, participants described familiar scenes of a patient reclining in a therapist's office. Geoff framed his description of therapy as knowledge ('what I know') based on television (perhaps implicitly indicating limitations of the accuracy of this information). The impact of media on knowledge of services may be buffered by the recognition that media portrayals are often inaccurate. However, with a lack of alternative sources of information, it may be difficult to differentiate realistic portrayals from inaccurate ones.

#### **'Maybe just deal with it yourself first'**

Participants described their preference for assessing and 'dealing' with their own problems before, and in some cases instead of, seeking help. The following extract gets to the heart of the masculine need for control. Kris has been asked to clarify his view on seeking counselling:

*Kris: Probably I'd hesitate to do it. It would be an uncomfortable thing to go and admit that you need help with something. That kind of shatters that fantasy that you can deal with everything yourself and move mountains. If it was something that needed to be done, it needs to be done.*

Kris highlights what is perhaps a typical masculine ideal of being able to fix problems and 'move mountains,' evoking a powerful sense of autonomy. Interestingly, he also describes a level of pragmatism which aligns with hegemonic masculinity - doing what 'needs to be done' despite discomfort. That is, despite his preference to fix problems himself, Kris acknowledges that he would seek help if necessary, which softens the rigidity of his allegiance to independence yet maintains his adherence to masculine values. In the following extract, Tane explains how he believes someone with a mental health problem could get help:

*Tane: Well I sort of have... I sort of have like a belief - maybe - it's not a really strong belief, like it's not set in stone, but um obviously trying to maybe just deal with it yourself first. Like try to have a look at yourself aye, like... Just try get in touch with yourself, see what you're doing - like who are you around, what are your habits, you know what I mean?*

For Tane, rather than accessing mental health services, the initial response to distress should be to attempt to assess what is causing the problem and then to fix the problem by changing habits and social groups. This response suggests controllability and agency in mental health. Redirecting his response to autonomous self-help, rather than explaining external help-seeking, may demonstrate the importance of independence. Notably, Tane stated that independent options should be attempted

'first' suggesting that external services may be acceptable if independent attempts fail.

When discussing talk-therapy, Manish explained his feelings on disclosing mental health problems:

*Manish: I'm... not really too comfortable doing that. I personally try to just handle it myself. If I can't do it, then yeah, I'll discuss it with my friends.*

Manish creates a hierarchy of steps which prioritises autonomously 'handling' the issue due to his discomfort in disclosing problems to others. Despite the context of discussing talk therapy, his next step would be to discuss the problems with his friends. Seeking professional services does not feature on his hierarchy and perhaps discussing the problem with friends deviates less from the imperative for autonomy than utilising professional services would.

#### **'So, it would be some form of social aspect to it.'**

When describing the process and benefits of talk-based therapy, participants tended to compare it to informal social support. Participants recognised the value of a confidential, safe space to talk; however, they also noted the limits of 'just talking':

*Timothy: I mean I think it's good. I think it is helpful. I think like - you know - while there might be people who have like a massive friend group in dealing with that stuff, they have every avenue in their own life to go and talk to anyone they want, there are people out there who are just - you know - they're in their home. They don't talk to their family, they don't really have friends, they don't feel they can trust anyone enough to talk to - I think in those kind of situations they are extremely helpful, but that being said, I don't think they're - kind of - the only thing that needs to happen - like some people just need a vent and they can get it out and they're fine, but obviously there are people who - they should be on medication. (...)*

Timothy acknowledged the usefulness of talk-therapy among people who lack trustworthy confidants to whom they can talk. However, he also outlined the limitations of 'venting,' noting that some people need medication. This suggests a hierarchy of mental distress and appropriate treatment, in which medication may be necessary for more severe mental distress.

Matt also described a hierarchy of steps needed for mental health support, noting that talking to friends and family were the first steps. He went on to explain the triggers for seeking the next step - seeking talk therapy:

*Matt: (...) If they (family) don't provide the environment for you to open up and get better then quickly seek the phone and then go from there. But if they provide an environment for you to open up and get help, then stay with them because you know they care and they - two ha- four hands is better than two hands. So, we have a support system and in mental health what you need most is a good support system to carry you through.*

To Matt, the most important part in overcoming mental health problems was having a 'good support system,' which should be provided by family, but may

also be provided by professional services if family are not providing the right environment to open up. In a similar example, Tane explains what he imagines would happen in talk therapy:

*Tane: (...) I would suspect there would be mainly – obviously talking about things on your mind that you're not - might not be comfortable talking to close family members with... which I'm not sure why you would be - maybe because you don't want to - you know, want them to see you.*

Tane appears to position talking to family members about mental health problems as the preferred strategy and therapy as an alternate. Likewise, for him it seems talk therapy adds value through providing a safe space to expose vulnerability when talking to friends or family appears too difficult or might make things worse. The fear of talking to family suggests a sense of self and expected stigma in exposing problems to family and could also allude to instances when the causes of problems are connected to family.

Although he avoids the comparison to family, Simon also points to the interactive social benefits of talk therapy when attempting to describe what he believes it involves:

*Simon: Um apart from like a touch base of one-on-one obviously, I think it would be more of um like touching base with - in terms of - hey look, emails and things like that. Or 'hey can I book another-' or 'look this has happened this week, I need to talk back and forth.' So, it would be some form of social aspect to it.*

Simon stated that a therapist would be helpful through their availability as a dedicated social support person. This description suggests that having someone available to talk to about day-to-day problems is useful to reduce mental health problems. Simon's description focuses on the social support provided by a therapist; however, he also alludes to the more formal aspects of the therapeutic relationship through mentioning the need to book another session.

#### **'You don't exactly know how to deal with coming off'**

Participants demonstrated a common set of beliefs that medication can lead to reliance. In this extract, Wiremu explains why he believes it is problematic for people to be inappropriately prescribed psychotropic medication:

*Wiremu: Probably prescription medication, I know you can become quite hooked on it, but also if you're on something for anti-depressants and you go on it for a really long time and then you have to come off for it, you don't exactly know how to deal with coming off of it.*

*Wiremu: (...) Like you go on it and then you're on it for - say - six months and then you're like 'okay no,' you're done and you decide to come off it, you're probably not gonna exactly know how to handle like your emotions or the moods that you'll suddenly be going through because you've just been so mellow and numb for like the past couple of months. Like I think it would be quite a bad shock for you. But then flowing on from that, like you may just turn to other ways of trying to cope*

*with that because you haven't dealt with your emotions over the past, say six months, because they've always just been like mellowed out, so you may just turn to other forms to try to cope with something - like alcohol or like other drugs or something, so...*

In describing the problems associated with weaning oneself off psychotropic medication, Wiremu created a scenario of potential problems for someone who has been on medication: reliance and withdrawal. The scenario suggests seeking emotional numbing, and possible substance abuse issues.

The following example illustrates the idea that participants saw prescription medication as leading to reliance, which may result in other negative outcomes. In this example, Arjun explained his beliefs regarding psychotropic medication:

*Arjun: I think that's just a start for another problem. Um I'm not anti-medicine but um I think that's just an opportunity for someone who's suffering through something to get addicted to something else. Something new which could lead to a bigger problem. I mean - people get addicted to Panadol - you know? So (laughs) it's - it's not that hard but I don't think it's the right thing to do to someone who's already mentally - not weak - but um you know - vulnerable in a way... you know - you're just giving them another reason to get addicted to something, which could - could later on turn into a worse problem you know. Cause once - once you stop giving them - it just leads... It's a spiral.*

Arjun described the sequelae of addiction, which could spiral into further complicating factors. He suggested that medication creates more problems for someone already struggling, which goes beyond other participants' suggestions that medication may result in reliance. Arjun also went to some effort not to equate weak with vulnerable, perhaps mitigating a negative gendered view of mental 'weakness' through his stress on a less negative 'vulnerable.'

#### **'Cause they're the doctor, they know what's best for you.'**

Despite describing discomfort with prescription medication and admitting the need for help, participants valued expert knowledge and advice of professional mental health care providers. Timothy would listen to his doctor if he were offered a prescription though he previously stated he felt uncomfortable using psychotropic medication:

*J: So, if you were to be prescribed one of these medications by someone, how would you feel about that? Would you take it?*

*Timothy: Yeah. I mean yeah - I would take it. Like my kind of view on any sort of medication is like 'I'm not a doctor,' like this person has gone through X amount of years of training to get to this point and like I'm not going to sit there and go like 'No! I'm not going to trust your judgement here because I read something on Facebook and they said medication's bad' so I'm guna - yeah I'm guna*

*trust the person who's got the personal experience talking to me and the - you know - the expertise of their degrees and qualifications and that kind of stuff. Yeah over just the random noise (laughs) of the world.*

As he explained his approach to psychotropic medication, Timothy created a dichotomy between expert knowledge and the 'random noise of the world,' which serves to dilute his previous expressions of concerns regarding medication, perhaps by classifying those beliefs among the 'random noise.' He acknowledged that many of his beliefs regarding medication were based on unreliable sources, which enabled him to act against his previously stated beliefs, and hypothetically accept the doctor's recommendation.

Prior to the following exchanges, Wiremu and Manish were discussing how they would respond to being offered a prescription medication. Fitting with the 'You don't exactly know how to deal with coming off' theme, both reported hesitancy; however, they softened that view by noting their trust in health professionals:

*(1) Wiremu: Yeah. But then again, it's one of those things where it's coming from your Doctor so you-I'd just actually think that what they're trying to give me would probably be best for me. Kind of thing, so it's one of those - 'Cause they're the doctor, they know - they know what's best for you, so it must be good for me.*

*(2) J: Okay and you don't really have much thought about how they actually work in terms of what they do.*

*Manish: Not really (laughs). Even now when I go to a doctor, if - for any illness or something, they just prescribe a medicine. If - sometimes I have time, I just go through and google the name and stuff, and like what it is, but at the end of the day I just take whatever's needed.*

These examples demonstrate a trust of the doctor for these participants; Manish and Wiremu indicated that the doctor's decision overrides their own attitudes and the doctor knows what is best for them, despite other hesitations. Manish does indicate some critical consumption of medical advice; however, this is sacrificed if under time pressure, in which case he would trust that the doctor has given him what he needs.

## DISCUSSION

We sought to improve understanding of young men's beliefs and attitudes towards mental health services, with consideration to how these beliefs may affect help-seeking. Five themes were formulated suggesting complex beliefs and attitudes towards services. This research contributed to the literature by exploring an area that has previously not been well-researched: the beliefs of men who have not accessed services. Addressing men's beliefs when they have no lived experience of services adds understanding to the important area of barriers to help-seeking and provides a stronger foundation to address such attitudinal barriers. By conducting an explorative, inductive study, themes relating to this area were able to be formulated, without a

reliance on existing hypotheses or deductive reasoning. This method was useful as it resulted in the generation of complex and rich data, allowing for more nuanced implications about the influences of knowledge, masculinity, and media on beliefs about services and possible impacts on help-seeking.

### Sources of knowledge

Participants reported that their knowledge about services was limited, mostly being sourced from television and movies. This finding aligns with the theses of Orchowski et al. (2006) and Wedding (2017) who both argued that film and television impact attitudes towards services. Likewise, McKelley and Rochlen (2007) suggested that men are unlikely to engage with talk-therapy due to a deficit of knowledge of talk-therapy, which is filled by negative portrayals in the media. Additionally, participants' description of images of 'comfy couches,' aligns with Harding and Fox's (2015) finding that men expected therapy to involve couches and suggests that both groups of men were influenced by similar fictional depictions of mental health services.

The finding that these men relied on media depictions for their descriptions of mental health services suggests a fundamental deficit in mental health literacy – an issue that previous research has suggested presents an important barrier to help seeking (Jorm 2012; Levant et al., 2009); however, the men in this research were somewhat more circumspect in their discussion of services. That is, despite describing ostensibly fearful beliefs of services as 'physical, aggressive... having some kind of traumatic experience,' (Arjun's words) and having cautious beliefs about seeking help 'I personally just try to handle it myself,' (Manish's words) participants also noted that they would seek help if they needed to. Indeed; while beliefs around harmful services, addictive medication, and masculine ideals about addressing problems without help may act as barriers, beliefs that services are socially supportive and that men should trust in experts are likely facilitators to help-seeking.

This study goes further than previous research regarding the impact of popular media on beliefs, as this research suggested awareness by participants that film and television portrayals may be inaccurate. Yet, despite awareness of the limitations of media portrayals, participants drew upon these representations in their descriptions of services. It appears that in the absence of alternative, more realistic depictions of services to draw upon, media depictions fill the void of understanding as McKelley and Rochlen (2007) suggested. However, it is also important to note that the extreme negative portrayals of services, which participants described, may reflect legitimate fears of services relating to higher rates of involuntary and harmful services experienced by Māori men (Drown, Harding, and Marshall, 2018) rather than resulting simply from popular media depictions.

### Masculine Attitudes towards help-seeking

Participants expressed a belief that services should be a final resort, and that they preferred to fix problems themselves first. This finding fits with literature regarding the impact of hegemonic masculinities on

help-seeking. That is, independence, control, and repressed emotionality are barriers that are incompatible with help-seeking (Cleary, 2012; Krum et al., 2017).

As well as expressing discomfort with the idea of seeking-help, participants expressed stoic pragmatism in their recognition that they would do what they must to improve their mental health. That is, participants noted that they might seek help, despite discomfort. This pragmatism also appears to reflect hegemonic masculine values – of taking control and enduring discomfort, but in a way that supports help-seeking. This finding challenges claims that hegemonic masculinity is inherently incompatible with help-seeking and adds nuance to understandings of the impacts of masculinity on help-seeking behaviours. A similar finding was reported by Ridge, Emslie, and White, (2011) who found that firefighters framed help-seeking as a masculine act of exerting control over their health.

### **Beliefs and attitudes towards specific services**

Without lived experience of talk-therapy, participants likened it to informal social supports such as talking with family and friends. Participants suggested that the primary benefit of talk-therapy was the provision of a safe and trustworthy space to discuss problems, particularly when informal supports were not providing this space. The finding that participants expressed that talk-therapy comprises supportive talking aligns with findings of Midgley et al. (2016), who found that young people expected therapy to be a chance to talk about problems. Based on Māori models of healthcare, where the importance of relationships is central (Hamley & Le Grice, 2021), this construction of talk-therapy may suggest positive beliefs towards talk-therapy from some participants.

It appeared that participants generally held positive beliefs regarding talk-therapy, consistent with literature indicating that men who *had* used services endorsed positive beliefs (Sierra et al., 2014). However, beliefs in the present study were mediated by participants' opinions as to how effective social support/talking was for coping with mental distress. That is, some participants described limitations of 'just talking,' fitting with other research indicating that men prefer medication to talk-therapy (Harris et al., 2016). Therefore, beliefs that talk-therapy comprises supportive talking may be a facilitator, or barrier to help-seeking, dependent on other factors such as the man's beliefs about the importance of talking and relationships.

Participants expressed concern about reliance on prescription medication. This contrasts with some literature suggesting men prefer medication (Harris et al., 2016). However, the results are equivocal as other studies have suggested that men prefer non-medical options (Sierra et al., 2014) and likewise, that the public in general perceives medication as addictive and unhelpful (Mirnezami et al., 2016). Participants' concern about medication broadly aligns with Harding and Fox's (2015) findings that men were concerned about medication before seeking treatment. However, the articulation of concerns differed as men in this study discussed concerns regarding reliance on medication, whereas participants in Harding and Fox's study

expressed concerns regarding the impact of medication on their personality.

It appeared that participants' disfavour of medication was contingent on the belief that medication leads to reliance and prevents long-term autonomous coping skills. Perhaps participants' preference to manage problems independently – likely connected to hegemonic masculine values of control and autonomy – relates to their preference not to use medication.

Participants expressed the belief that mental health professionals know better than laypeople, and that their advice should be trusted. In some instances, participants even noted that they would accept their doctor's advice when it goes against other beliefs they hold. This apparent discrepancy may reflect that negative attitudes can be attenuated by the normativeness of a situation (that is, it is *normal* to listen to a doctor), and that more credible sources of information (such as doctors) can have stronger effects on beliefs than less credible sources (Fazio, 1986; Maio et al., 2018). This finding presents an interesting conflict between situationally normative behaviours and broader hegemonic masculine behaviours. Pattyn et al. (2015) articulated this conflict by highlighting the strains of maintaining the role of both a masculine man and a patient. This finding again highlights the complexity of attempting to understand the impact of masculine roles on help-seeking.

### **Research Applications**

Based on the findings of this study, improving education regarding mental health services is warranted. Previous research has shown that education campaigns targeted to men reduce barriers to help-seeking through improving attitudes towards services and reducing fear of stigma (Hammer & Vogel, 2010). Additionally, men have expressed that education regarding mental health and services would be useful (e.g., Harding & Fox, 2015; McKelley & Rochlen, 2007). Substantial research has found that health literacy can be a barrier to help-seeking (Olliffe et al., 2020), and this study suggests deficits in mental health literacy among young men. In addition, growing research indicates that educational programmes designed for men, and which relate to key issues may be effective in improving literacy and help-seeking (Olliffe et al., 2020). Therefore, the development (and evaluation) of programmes designed for young men, which address beliefs highlighted in this study (such as the belief that services involve coercive and forceful treatment, and that medication is addictive) may be an effective application of the findings of this study. Likewise, providing choice in treatment options appears to be important in facilitating help-seeking, as the young men in this study expressed various preferences and opinions.

Education campaigns regarding the content of services may address several topics based on this research. It may be useful to have short 'profile' type videos, in which service providers describe who they are and their role. This would serve to offer more alternative sources of information to film and television. These videos may also focus on specific barriers highlighted in this study – for example, a talk-therapist explaining what talk therapy involves, and how it goes beyond 'just

talking.' Likewise, a medication prescribing clinician may openly discuss the possible risks and benefits of medication, with a specific focus on the fear that medication leads to reliance. Additionally, it appears that each of these videos should focus on framing services as supporting people to be independent and giving them tools to manage their distress (to align with the preference to fix problems independently). Such education campaigns may be disseminated on television and social media, which have been shown as effective platforms for public health campaigns in international and New Zealand indigenous Māori contexts (Austin et al., 2015; Wilson et al., 2005).

### Limitations and future research

Despite the demographic diversity in this study, it is possible that through voluntary participation, this study included young men interested in mental health. It may also have been useful to ensure participation of men who identified as transgender, in addition to cisgender men, to further increase the diversity of the participants. Likewise, it is possible that the process of discussing and articulating mental health services in interviews may have facilitated the construction of more nuanced, explicit beliefs towards services (like Tomm's (1987) notion that questions are never neutral). That is, the construction of beliefs in such a study as this are likely different to the construction of beliefs in other situations that may be more relevant to 'real life' help-seeking.

This study is premised on the inferred link between opinions of young men and their behaviours in relation to help-seeking. Although research does support such a link between attitudes and behaviours (Petty, 2018), it is well established that there are limitations to this link. It is also worth noting that some preferences expressed in this study, such as the preference to fix problems independently may reflect common, non-gendered, responses to health problems such as normalising of problems (Biddle et al., 2007), and avoidance of the time and cost commitment of utilising services (Andrade et al., 2014). Nevertheless, these preferences were expressed by young men and tended to reflect hegemonic pressures for independence and control.

### References

- Andrade, L. H., Alonso, J., Mneimneh, Z., Wells, J., Al-Hamzawi, A., Borges, G., . . . De Graaf, R. (2014). Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychological Medicine*, 44(6), 1303-1317. <https://doi.org/10.1017/S0033291713001943>
- Austin, A., Barnard, J., Hutcheon, N., & Parry, D. (2015). Media consumption forecasts 2015. *Режим доступу: https://communicateonline.me/wpcontent/uploads/2016/06/Media-Consumption-Forecasts-2016.pdf*
- Biddle, L., Donovan, J., Sharp, D., & Gunnell, D. (2007). Explaining non-help-seeking amongst young adults with mental distress: a dynamic interpretive model of illness behaviour. *Sociology of Health & Illness*, 29(7), 983-1002. <https://doi.org/10.1111/j.1467-9566.2007.01030.x>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>

Future research may build upon these findings by interviewing more specific groups of men, such as men who identified as experiencing mental distress but have not sought help and groups of men with cultures that value different models of healthcare. Such studies would allow for a more deductive exploration of beliefs that contributed to perceived barriers to help-seeking. Likewise, men living in areas of high deprivation may be interviewed similarly to the method used in the present study. Higher rates of deprivation have been found to have a negative impact on help-seeking (Keown et al., 2016); however, deprivation was not an inclusion criterion in the present study. Future research may also use more deductive questioning methods to explore specific 'myths' about services, based on the foundation of this, and other studies.

### Conclusions

This qualitative exploratory study explored the beliefs and attitudes of young men who had never accessed mental health services. The findings indicate that these young men lacked information to form confident opinions of services and relied on various indirect and possibly unreliable sources, including popular media, to inform their opinions. Although they preferred to maintain their autonomy and control over their health decisions, several were still willing to seek treatment despite expressing such values. Furthermore, various non-gendered belief systems such as adhering to the doctor's advice and valuing a support network were apparent.

These findings indicate that service providers and governmental agencies need to go further in addressing attitudinal barriers to help-seeking amongst young men, while researchers may further explore the form and impact of the beliefs identified in this study. Although hegemonic masculine values inform beliefs and attitudes towards help-seeking, this study suggests greater level of nuance and complexity at work in these young men's considerations around service utilisation and provides hope that young men are critical consumers of knowledge and may be open to information that softens attitudinal barriers to help-seeking.

- Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological* (p. 57-71). American Psychological Association. <https://doi.org/10.1037/13620-004>
- Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic analysis. *Handbook of Research Methods in Health Social Sciences*, 843-860. Springer Singapore.
- Burnard, P., Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Analysing and presenting qualitative data. *British Dental Journal*, 204(8), 429. <https://doi.org/10.1038/sj.bdj.2008.292>
- Burr, V. (2015). *Social Constructionism*: Routledge.
- Cleary, A. (2012). Suicidal action, emotional expression, and the performance of masculinities. *Social Science & Medicine*, 74(4), 498-505. <https://doi.org/10.1016/j.socscimed.2011.08.002>

- Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., . . . Thornicroft, G. (2015). What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychological Medicine*, 45(1), 11-27. <https://doi.org/10.1017/S0033291714000129>
- Coles, R., Watkins, F., Swami, V., Jones, S., Woolf, S., & Stanistreet, D. (2010). What men really want: a qualitative investigation of men's health needs from the Halton and St Helens Primary Care Trust men's health promotion project. *British Journal of Health Psychology*, 15(4), 921-939. <https://doi.org/10.1348/135910710X494583>
- Cramer, K. M. (1999). Psychological antecedents to help-seeking behavior: A reanalysis using path modeling structures. *Journal of Counseling Psychology*, 46(3), 381. <https://doi.org/10.1037/0022-0167.46.3.381>
- Drown, C., Harding, T., & Marshall, R. (2018). Nurse perceptions of the use of seclusion in mental health inpatient facilities: have attitudes to Māori changed?. *The Journal of Mental Health Training, Education and Practice*, 13(2), 100-111. <https://doi.org/10.1108/JMHTEP-12-2016-0055>
- Fazio, R. H. (1986). How do attitudes guide behavior. *Handbook of Motivation and Cognition: Foundations of Social Behavior*, 1, 204-243.
- Furnham, A. (2009). Psychiatric and psychotherapeutic literacy: attitudes to, and knowledge of, psychotherapy. *International Journal of Social Psychiatry*, 55(6), 525-537. <https://doi.org/10.1177/0020764008094428>
- Furnham, A., & Telford, K. (2012). Public attitudes, lay theories and mental health literacy: The understanding of mental health. In *Mental Illnesses-Understanding, Prediction and Control*: InTech.
- Gergen, K. J. (1994). Exploring the postmodern: Perils or potentials? *American Psychologist*, 49(5), 412. <https://doi.org/10.1037/0003-066X.49.5.412>
- Gergen, K. J. (2009). *Realities and Relationships: Soundings in Social Construction*: Harvard university press.
- Gonzalez, J. M., Alegria, M., & Prihoda, T. J. (2005). How do attitudes toward mental health treatment vary by age, gender, and ethnicity/race in young adults?. *Journal of community psychology*, 33(5), 611-629. <https://doi.org/10.1002/jcop.20071>
- Hamley, L., & Grice, J. L. (2021). He kākano ahau—identity, Indigeneity and wellbeing for young Māori (Indigenous) men in Aotearoa/New Zealand. *Feminism & Psychology*, 31(1), 62-80. <https://doi.org/10.1177/0959353520973568>
- Hammer, J. H., & Vogel, D. L. (2010). Men's help seeking for depression: The efficacy of a male-sensitive brochure about counseling. *The Counseling Psychologist*, 38(2), 296-313.
- Harding, C., & Fox, C. (2015). It's not about "Freudian couches and personality changing drugs" An investigation into men's mental health help-seeking enablers. *American Journal of Men's Health*, 9(6), 451-463. <https://doi.org/10.1177/1557988314550194>
- Harris, M., Baxter, A., Reavley, N., Diminic, S., Pirkis, J., & Whiteford, H. (2016). Gender-related patterns and determinants of recent help-seeking for past-year affective, anxiety and substance use disorders: findings from a national epidemiological survey. *Epidemiology and Psychiatric Sciences*, 25(6), 548-561. <https://doi.org/10.1017/S2045796015000876>
- House, J., Marasli, P., Lister, M., & Brown, J. S. (2018). Male views on help-seeking for depression: A methodology study. *Psychology and Psychotherapy: Theory, Research and Practice*, 91(1), 117-140. <https://doi.org/10.1111/papt.12144>
- Hudson, M.L., Russell, K. (2009). The Treaty of Waitangi and Research Ethics in Aotearoa. *Bioethical Inquiry*, 6, 61-68. <https://doi.org/10.1007/s11673-008-9127-0>
- Jewkes, R., Morrell, R., Hearn, J., Lundqvist, E., Blackbeard, D., Lindegger, G., ... & Gottzén, L. (2015). Hegemonic masculinity: combining theory and practice in gender interventions. *Culture, health & sexuality*, 17(2), 112-127. <https://doi.org/10.1080/13691058.2015.1085094>
- Jorm, A. F. (2000). Mental health literacy: public knowledge and beliefs about mental disorders. *The British Journal of Psychiatry*, 177(5), 396-401. <https://doi.org/10.1192/bjp.177.5.396>
- Jorm, A. F. (2012). Mental health literacy: Empowering the community to take action for better mental health. *American Psychologist*, 67(3), 231. <https://doi.org/10.1037/a0025957>
- Keown, P., McBride, O., Twigg, L., Crepaz-Keay, D., Cyhlarova, E., Parsons, H., . . . Weich, S. (2016). Rates of voluntary and compulsory psychiatric in-patient treatment in England: an ecological study investigating associations with deprivation and demographics. *The British Journal of Psychiatry*, 209(2), 157-161. <https://doi.org/10.1192/bjp.bp.115.171009>
- Kerridge, B. T., Pickering, R. P., Saha, T. D., Ruan, W. J., Chou, S. P., Zhang, H., . . . Hasin, D. S. (2017). Prevalence, sociodemographic correlates and DSM-5 substance use disorders and other psychiatric disorders among sexual minorities in the United States. *Drug and Alcohol Dependence*, 170, 82-92. <https://doi.org/10.1016/j.drugalcdep.2016.10.038>
- Kessler, D., Heath, I., Lloyd, K., Lewis, G., & Gray, D. P. (1999). General Practice Cross sectional study of symptom attribution and recognition of depression and anxiety in primary care Commentary: There must be limits to the medicalisation of human distress. *BMJ*, 318(7181), 436-440. <https://doi.org/10.1136/bmj.318.7181.436>
- Krumm, S., Checchia, C., Koesters, M., Kilian, R., & Becker, T. (2017). Men's views on depression: a systematic review and metasynthesis of qualitative research. *Psychopathology*, 50(2), 107-124. <https://doi.org/10.1159/000455256>
- Lauber, C., Carlos, N., & Wulf, R. (2005). Lay beliefs about treatments for people with mental illness and their implications for antistigma strategies. *The Canadian Journal of Psychiatry*, 50(12), 745-752. <https://doi.org/10.1177/070674370505001203>
- Levant, R. F., Hall, R. J., Williams, C. M., & Hasan, N. T. (2009). Gender differences in alexithymia. *Psychology of Men & Masculinity*, 10(3), 190. <https://doi.org/10.1037/a0015652>
- Lynch, L., Long, M., & Moorhead, A. (2018). Young men, help-seeking, and mental health services: exploring barriers and solutions. *American journal of men's health*, 12(1), 138-149. <https://doi.org/10.1177/1557988315619469>
- Maio, G. R., Haddock, G., & Verplanken, B. (2018). *The Psychology of Attitudes and Attitude Change*: Sage Publications Limited.
- McKelley, R. A., & Rochlen, A. B. (2007). The practice of coaching: Exploring alternatives to therapy for counseling-resistant men. *Psychology of Men & Masculinity*, 8(1), 53. <https://doi.org/10.1037/1524-9220.8.1.53>



- Midgley, N., Holmes, J., Parkinson, S., Stapley, E., Eatough, V., & Target, M. (2016). "Just like talking to someone about like shit in your life and stuff, and they help you": Hopes and expectations for therapy among depressed adolescents. *Psychotherapy Research*, 26(1), 11-21. <https://doi.org/10.1080/10503307.2014.973922>
- Mirnezami, H. F., Jacobsson, L., & Edin-Liljegren, A. (2016). Changes in attitudes towards mental disorders and psychiatric treatment 1976–2014 in a Swedish population. *Nordic Journal of Psychiatry*, 70(1), 38-44. <https://doi.org/10.3109/08039488.2015.1046916>
- Möller-Leimkühler, A. M. (2002). Barriers to help-seeking by men: a review of sociocultural and clinical literature with particular reference to depression. *Journal of Affective Disorders*, 71(1-3), 1-9. [https://doi.org/10.1016/S0165-0327\(01\)00379-2](https://doi.org/10.1016/S0165-0327(01)00379-2)
- New Zealand Government. Ministry of Health. (2016). *Suicide Facts: Data tables 1996–2016*. Published online by the Ministry of Health. <https://www.health.govt.nz/publication/suicide-facts-data-tables-19962016>
- Oliffe, J. L., Rossnagel, E., Kelly, M. T., Bottorff, J. L., Seaton, C., Darroch, F., Men's health literacy: a review and recommendations, *Health Promotion International*, 35(5), October 2020, Pages 1037-1051, <https://doi.org/10.1093/heapro/daz077>
- Orchowski, L. M., Spickard, B. A., & McNamara, J. R. (2006). Cinema and the valuing of psychotherapy: Implications for clinical practice. *Professional Psychology: Research and Practice*, 37(5), 506. <https://doi.org/10.1037/0735-7028.37.5.506>
- Pattyn, E., Verhaeghe, M., & Bracke, P. (2015). The gender gap in mental health service use. *Social Psychiatry and Psychiatric Epidemiology*, 50(7), 1089-1095. <https://doi.org/10.1007/s00127-015-1038-x>
- Petty, R. E. (2018). *Attitudes and Persuasion: Classic and Contemporary Approaches*: Routledge.
- Ramaeker, J., & Petrie, T. A. (2019). "Man up!": Exploring intersections of sport participation, masculinity, psychological distress, and help-seeking attitudes and intentions. *Psychology of Men & Masculinities*, 20(4), 515. <https://doi.org/10.1037/men0000198>
- Rice, S. M., Telford, N. R., Rickwood, D. J., & Parker, A. G. (2018). Young men's access to community-based mental health care: qualitative analysis of barriers and facilitators. *Journal of Mental Health*, 27(1), 59-65. <https://doi.org/10.1080/09638237.2016.1276528>
- Ridge, D., Emslie, C., & White, A. (2011). Understanding how men experience, express and cope with mental distress: where next? *Sociology of Health & Illness*, 33(1), 145-159. <https://doi.org/10.1111/j.1467-9566.2010.01266.x>
- Schultz, H. (2005). Good and bad movie therapy with good and bad outcomes. *The Amplifier*, 19.
- Seidler, Z. E., Dawes, A. J., Rice, S. M., Oliffe, J. L., & Dhillon, H. M. (2016). The role of masculinity in men's help-seeking for depression: a systematic review. *Clinical psychology review*, 49, 106-118. <https://doi.org/10.1016/j.cpr.2016.09.002>
- Sierra Hernandez, C. A., Oliffe, J. L., Joyce, A. S., Söchting, I., & Ogrodniczuk, J. S. (2014). Treatment preferences among men attending outpatient psychiatric services. *Journal of Mental Health*, 23(2), 83-87. <https://doi.org/10.3109/09638237.2013.869573>
- Swami, V. (2012). Mental health literacy of depression: gender differences and attitudinal antecedents in a representative British sample. *PLoS one*, 7(11), e49779. <https://doi.org/10.1371/journal.pone.0049779>
- Tan, Y., Shaw, P., Cheng, H., & Kim, K. K. (2013). The construction of masculinity: A cross-cultural analysis of men's lifestyle magazine advertisements. *Sex Roles*, 69(5-6), 237-249. <https://doi.org/10.1007/s11199-013-0300-5>
- Tuffin, K., & Danks, J. (1999). Community care and mental disorder: An analysis of discursive resources. *British Journal of Social Psychology*, 38(3), 289-302. <https://doi.org/10.1348/014466699164176>
- Vogel, D. L., Gentile, D. A., & Kaplan, S. A. (2008). The influence of television on willingness to seek therapy. *Journal of Clinical Psychology*, 64(3), 276-295. <https://doi.org/10.1002/jclp.20446>
- Vogel, D. L., Heimerdinger-Edwards, S. R., Hammer, J. H., & Hubbard, A. (2011). "Boys don't cry": Examination of the links between endorsement of masculine norms, self-stigma, and help-seeking attitudes for men from diverse backgrounds. *Journal of Counseling Psychology*, 58(3), 368. <https://doi.org/10.1037/a0023688>
- Walker, E. R., Cummings, J. R., Hockenberry, J. M., & Druss, B. G. (2015). Insurance status, use of mental health services, and unmet need for mental health care in the United States. *Psychiatric Services*, 66(6), 578-584. <https://doi.org/10.1176/appi.ps.201400248>
- Wardell, S. E. (2013). Doctors and All Blacks: How depression and its treatment is framed in New Zealand GP-targeted advertising. *Sites: a Journal of Social Anthropology and Cultural Studies*, 10(2), 52-81. <https://doi.org/10.11157/sites-vol10iss2id217>
- Wedding, D. (2017). Public education and media relations in psychology. *The American Psychologist*, 72(8), 764-777. <https://doi.org/10.1037/amp0000202>
- Wilson, N., Grigg, M., Graham, L., & Cameron, G. (2005). The effectiveness of television advertising campaigns on generating calls to a national Quitline by Māori. *Tobacco Control*, 14(4), 284-286. <http://dx.doi.org/10.1136/tc.2004.010009>
- World Health Organization. (2002). *The World Health Report 2002: Reducing Risks, Promoting Healthy Life*. Published online by the World Health Organization. <https://www.who.int/whr/2002/en/>
- Yousaf, O., Grunfeld, E. A., & Hunter, M. S. (2015). A systematic review of the factors associated with delays in medical and psychological help-seeking among men. *Health Psychology Review*, 9(2), 264-276. <https://doi.org/10.1080/17437199.2013.840954>

### Corresponding Author

Jake Gallagher  
 Massey University Wellington  
 PO Box 756, Wellington 6140  
 Email: [Jakemgallagher92@gmail.com](mailto:Jakemgallagher92@gmail.com)  
 Phone: +64 277746653

## Age stratified normative data for Māori on the Wechsler Adult Intelligence Scale (4th edition; WAIS-IV)

Makarena Dudley<sup>1</sup>, Suzanne Barker-Collo<sup>1</sup>, Denise Wilson<sup>2</sup>, and Nick Garrett<sup>3</sup>

<sup>1</sup>School of Psychology, The University of Auckland, New Zealand

<sup>2</sup>Tapua Wairoa Māori Health Research Centre, Auckland University of Technology

<sup>3</sup>Biostatistics and Epidemiology, Auckland University of Technology

---

Culture impacts neuropsychological test performance. This has been shown to be the case for Māori, the Indigenous people of New Zealand. The current study presents normative data that reflects a normative level of performance for a Māori population on the Wechsler Adult Intelligence Scale (4th edition; WAIS-IV). Participants were 284 neurologically healthy adults who self-identified as Māori, stratified across gender and seven age ranges, were recruited from seven different areas of the North and South Islands of New Zealand and were representative of the main Māori iwi/tribes. They were administered the WAIS-IV according to standard criteria. Normative data are presented across subtests for each of the seven age ranges. Normative data are not presented by gender as an ANOVA indicates few significant differences by gender.

**Keywords:** *Māori, neuropsychology, assessment, normative data, Wechsler Adult Intelligence Scale*

### INTRODUCTION

It is now widely acknowledged that culture impacts neuropsychological test performance (Ardila, 2007; Franzen et al., 2021; Pedraza & Mungas, 2008; Uzzellet al., 2013). Individuals from culturally and linguistically diverse backgrounds may be disadvantaged on neuropsychological tests, which are predominantly developed in western countries. This is of relevance in an Aotearoa (New Zealand) context where disparities are already evident in the incidence and health outcomes of neurological conditions for Māori, the Indigenous people of Aotearoa.

For example, the average age of stroke onset for Māori is 61 years, compared to 64 years for Pasifika people and over 75 years for Pākehā (European New Zealanders (Feigin et al., 2006). There is also some evidence that the chance of being dependent at 12 months post-stroke is three times higher for Māori compared to Pākehā (McNaughton et al., 2002; Ministry of Health, 2003). Furthermore, for traumatic brain injury (TBI), compared with Pākehā, Māori have a greater risk of mild TBI (RR 1.23, 95% CI 1.08-1.39), accounting for 31% of all TBIs despite comprising only 16.5% of the population (NZ Statistics, 2018).

These disparities persist despite the Ministry of Health's commitment to fulfil the special relationship between Māori and the Crown under Te Tiriti o Waitangi (Ministry of Health CO, 2019). This commitment demands response to the widespread and ongoing impact of colonisation that deeply harms Māori and continues to negatively affect the health of Māori (Ministry of Health, 2019).

Neuropsychological testing is commonly administered to individuals who have experienced brain injury as part of their assessment and treatment plan. However, neuropsychological tests that are widely used in Aotearoa have almost entirely been developed in either the United States or the United Kingdom and, as such, the test developers, the test content, and the standardised data that accompany the tests reflect a western worldview and tend to disadvantage individuals who diverge from a western cultural background (Manly, 2005; Tan et al., 2021). Only a few studies have investigated the cultural bias in neuropsychological measures when applied to Māori (Dudley et al., 2017; Ogden et al., 2003; Ogden & McFarlane-Nathan, 1997; Shepherd, & Leathem, 1999; Zawaly, et al., 2019), with only one study adopting a Kaupapa Māori Methods approach (Haitana et al., 2010). Collectively, these studies provide cumulative evidence of test bias in neuropsychological testing when assessing Māori.

Test bias can manifest in several ways including, but not limited to, construct bias, method bias and item bias. Construct bias is present when the concept being measured is not equivalent across cultural groups. Method bias occurs when variations in responses are caused by the instrument rather than the actual predispositions of the respondents that the instrument is attempting to uncover, and item bias is the presence of some characteristic of an item that results in differential performance for individuals of the same ability but from different ethnic, sex, cultural, or religious groups. (Fernández, & Abe, 2018; Pedraza, 2020; van de Vijver, & Tanzer, 2004).

Anastasi and Urbina (1997) also argue that all neuropsychological tests and accompanying normative data favour people from the same culture as the test



developers and so, to conduct an unbiased assessment, it is important that normative data is appropriate for the client assessed. In Aotearoa, a survey of psychologists and neuropsychologists was conducted to determine which factors influence test selection (Ross-McAlpine et al., 2018). As well as the primary question 66% of the sample also voiced concerns about the cultural sensitivity of some tests and thought that New Zealand normative data was needed (Ross-McAlpine et al., 2018).

Extant research suggests that Māori perform better on measures that have been adapted to include content that is familiar to them. For example, Ogden and McFarlane (1997), and Ogden et al. (2003) found that the performance of the Māori participants improved when the test items were adapted to reflect a Māori world view. Conversely, in their evaluation of the Peabody Picture Vocabulary Test-111 with Māori children, Haitana et al. (2010), found that some of the variance in the overall lower performance of Māori children was due to a lack of exposure to English words by those who attended Māori-medium schools. Familiarity with test content has been identified by cross-cultural neuropsychologists as an advantage with test takers (Manly, 2005).

Ogden and McFarlane (1997) also found that when responding to the Vocabulary subtest of the WAIS-R, the Māori males in their study provided answers that differed from the standard scoring criteria as prescribed in the WAIS-R manual, and therefore, received no points despite their answers being familiar to them and arguably 'correct' in their vernacular. These authors also found cultural differences in the perceived level of difficulty of certain stimuli, suggesting that knowledge that is standard or valued by Māori, may differ from conventional material against which they were being evaluated.

In addition to test content being a source of bias, evidence also shows that the neuropsychological process itself can impact on the performance of individuals who belong to ethnic minority groups (Brickman, et al., 2006). Factors such as the assessment setting, having a culturally matched assessor, the cultural competence of the clinician, the attitude of the examinee toward tests, as well as heightened anxiety of the person being assessed may have an influence on performance. Studies conducted in Aotearoa with Māori report similar findings (Dudley et al 2019; Ogden & McFarlane, 1997; Ogden et al. 2003; Shepherd & Leatham, 1999).

In the international literature methodologies to improve the reliability and validity of neuropsychological measures when used in diverse populations have been suggested by various cross-cultural neuropsychologists and include but are not limited to; the modification or discontinuation of tests that are not salient or relevant to a particular culture or language; to construct tests which are more culture fair and salient to diverse cultural groups, or; obtain local normative data appropriate for specific groups; Feigin and Barker-Collo (2007) have argued for some years for local normative data in Aotearoa. The current study has chosen to address the existing situation by obtaining data that reflects a normative level of performance for a

Māori population on the Wechsler Adult Intelligence Scale subtests (Wechsler, 2008).

The WAIS-IV was selected as the focus of this research as it is the most widely used neuropsychological measure for assessing cognitive functioning in adults 16 years and older, in Aotearoa. However, the standardised normative data that accompanies the WAIS-IV were derived from an American population and are the yardstick to which Māori who undergo this assessment are compared. Our previous publication (Dudley et al., 2019) provides means and standard deviations stratified for age and gender in a Maori sample for this measure. The study found that variables such as income and education may be factors that impact the performance of Māori although gender was not found to impact performance. The present study builds upon these previous findings by presenting normative conversion tables which allow clinicians to convert raw scores to their scaled score equivalents for each WAIS-IV subtest for each age range.

## METHOD

This study aligned with some of the domains and criteria provided by the CONSIDER statement for the reporting of research that aims to strengthen Indigenous health research and advance Indigenous health outcomes and development (Huria et al., 2019). A Māori-centred approach was adopted for this study in that the whole sample which was comprised of participants who identified as Māori, were administered an assessment tool from a western knowledge base. The data were analysed using the SPSS Data Analysis Software.

### Participants

Participants were 284 neurologically healthy adults who self-identified as Māori. Potential participants were screened for conditions such as major depressive disorder that could possibly affect cognitive test performance. Exclusion criteria were the same as those used for the WAIS-IV standardisation sample (refer to the WAIS-IV Technical and Interpretive Manual). Participation in the study was voluntary and participants provided written consent. All participants spoke English fluently, which reflected the findings of the New Zealand census whereby 96.1% of the Aotearoa population are English fluent (Statistics New Zealand, 2013). The sample was stratified for gender and grouped into seven age brackets (16-20 years, 21-30 years, 31-40 years, 41-50 years, 51-60 years, 61-70 years, 70+ years). The age range was from 16 years to 90 years ( $M = 45.40$ ,  $SD = 19.84$ ). Years of education was grouped to approximate grades of the Aotearoa education system with most of the sample having completed some high school or obtained a tertiary qualification.

Recruitment occurred across seven locations throughout the North and South islands of Aotearoa to maximise representation of the major iwi (tribes) throughout the country. Those iwi included: Ngāpuhi, Te Rarawa, Te Aupōuri, Ngāti Kūri, Ngāti Hine, Ngāti Whātua, Tainui, Tuhoe, Ngāti Maniapoto, Ngāti Tūwharetoa, Ngāti Porou, Whānau-ā-Apanui, Ngāti Kahungunu, Ngai Tahu, Te Arawa, Ngāti Awa, Te Ati Haunui-ā-Pāpārangi, Ngāti Raukawa, and Ngāti Tama. Most of the sample affiliated to one iwi (66%), whilst

22% identified with 2 iwi, 7% with 3 iwi, and 5% with 4 or more iwi. The number of participants recruited from the North and South Islands were proportionate to the total Māori population for each island as indicated in the New Zealand 2013 Census (i.e., 90% of Māori live in the North Island; (Statistics New Zealand, 2013). A summary of the demographic data is presented in Table 1.

**Table 1.** Characteristics of the sample

Characteristic	N (total = 284)	%
<b>Gender</b>		
Male	140	49.3
Female	144	50.7
<b>Age (years)</b>		
16-20	40	14.1
21-30	41	14.4
31-40	43	15.1
41-50	40	14.1
51-60	40	14.1
61-70	40	14.1
71+	40	14.1
<b>Education Years (completed)</b>		
≤5 (primary school)	5	1.8
6-7 (intermediate school)	12	4.2
8-12 (high school)	137	48.2
≥ 13 (tertiary)	130	45.8
<b>Household Income</b>		
\$0 - \$10,000	32	11.3
\$11,000 - \$20,000	77	27.1
\$21,000 - \$30,000	46	16.2
\$31,000 - \$40,000	28	9.9
\$41,000 - \$50,000	33	11.6
\$51,000 - \$60,000	5	1.8
\$61,000 - \$70,000	23	8.1
\$71,000 +	40	14.1

**The Wechsler Adult Intelligence Scale-IV**

The WAIS-IV (Wechsler, 2008) is a battery of tasks assessing various aspects of cognition. The battery contains 15 subtests, with raw scores on each subtest converted to scaled scores using normative data tables. Each subtest scaled score has a mean of 10 and standard deviation of 3.

Ten of the 15 subtests produce composite scores: Full Scale IQ (FSIQ), Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), and Processing Speed Index (PSI) based on age-corrected scaled scores. All 15 subtests were administered and scored in accordance with standardised procedures. The WAIS-IV Index/IQ are often described qualitatively that characterises the examinee’s level of composite score performance relative to same-age peers. Qualitative ranges include: 130 and above = Very Superior, 120 - 129 = Superior, 110 – 119 = High Average, 90 – 109 = Average, 75 – 89 = Low Average, 70 – 79 = Borderline, 69 and below = Extremely Low (WAIS-IV Technical and Interpretive Manual, p 126).

**Procedure**

The study was approved by the Auckland University of Technology Ethics Committee. Flyers outlining the study were distributed at universities and Māori health clinics throughout Aotearoa. The study was advertised on Māori radio stations and live presentations were delivered to Māori community groups at all seven

recruitment sites. Recruitment, however, was also achieved through whakawhanaungatanga (connections) utilising the researchers’ extensive networks within the Māori community. Recruitment was completed over a period of 18 months.

Once someone was identified as a possible participant, they were contacted by phone or face-to-face to ascertain their eligibility for the study. Those who met eligibility criteria and who provided verbal consent were then scheduled a time and place to conduct the assessment. Each participant was given a choice to hold the assessment at their home or another place of their choice such as their marae (Iwi meeting house), or at a Māori-friendly hauora organisation such as the Ahipara Medical Clinic in Te Tai Tokerau, and He Waka Tapu in Ōtautahi, or a Māori-friendly research clinic such as Te Atawhai Ō Te Ao in Wanganui. Tikanga guided the interview. Karakia (prayer), pepeha (introductions), whanaungatanga (the process of making connections) and kai (food) were all protocols that were observed. Te Reo Māori (the Māori language) was spoken when appropriate. The Participant Information Sheet was read through with the participant to clarify areas of uncertainty and to provide the participant with the opportunity to ask pātai (questions). Those who wished to continue with the interview provided written informed consent. Most administrations took place in a Māori-friendly research clinic (Māori NGO office; n = 224) office workplaces (n = 7), mainstream health clinics (n = 9), marae (n = 3), or at the participant’s home (n = 37). All administrative settings conformed to the guidelines for physical environment as stated in the WAIS-IV, Administration and Scoring Manual.

All measures were administered by either the first author who is Māori or a Māori research assistant who held a background in health at a tertiary level. The research assistants were extensively trained in the administration of the WAIS-IV. The time taken to complete the assessment ranged from 2½ hours to 4 hours. Participants were informed they could break for a rest whenever they felt they needed to. Once the WAIS-IV had been administered and completed each participant was thanked and given a koha (gift) as a token of appreciation for their participation.

Quality assurance measures adopted included contacting random participants by phone and enquiring about their experience and to determine if they had made their best effort. Random checks of 10% of participant score sheets were also conducted to ensure accuracy of scoring and of data entry.

Means and standard deviations were generated for each subtest and each age range. Normative data tables were then generated for each age range which allows conversion of raw scores on each subtest to scaled scores with a mean of 10 and standard deviation of 3.

The privacy of the participants was maintained by the de-identification of their individual score sheets which was replaced with a code. The data was entered onto a password protected spreadsheet which was accessible only to the researchers involved in the study.

**RESULTS**

Table 2 presents raw score means and standard deviations for each age range on each of the 15 WAIS-

IV subtests. A one-way ANOVA with gender as the grouping variable and raw score performance on each of the WAIS-IV subtests indicated that significant between group differences were present for only two subtests, both from the Processing Speed Index (i.e., Digit Symbol, Coding and Symbol Search), with males producing better performances. As such, the normative data are presented here by age range, but not separately by gender.

Tables 3 through 9 present conversion tables, which can be used to convert raw scores on WAIS-IV subtest to

scaled scores with a mean of 10 and standard deviation of 3. Each table presents data for a different age range. Norms were calculated for each age range and each subtest separately using the mean score and standard deviation of that score and then fitting the data to a normal distribution. For ease of clinical application, the format of the tables has been designed to replicate those currently used by clinicians from the WAIS-IV manual.

**Table 2.** Means and standard deviations of raw scores for each subtest across age groups

		16-20 years	21-30 years	31-40 years	41-50 years	51-60 years	61-70 years	71+ years
Block Design	Mean	48.5	49.98	47.47	47.05	39.33	34.45	30.65
	SD	9.89	9.31	8.96	9.33	8.81	9.00	10.55
Similarities	Mean	22.63	23.54	25.42	25.88	25.53	23.38	22.45
	SD	3.85	5.84	4.12	4.01	5.22	3.99	4.50
Digit span	Mean	29.3	29.73	30.28	30.35	28.55	26.35	23.55
	SD	4.76	4.18	5.50	5.01	4.35	4.43	4.43
Matrix Reasoning	Mean	19.08	18.88	16.98	18.20	14.53	12.30	10.95
	SD	3.87	3.78	4.94	4.11	4.90	4.63	4.48
Vocabulary	Mean	24.56	30.20	35.93	38.20	39.80	35.28	33.20
	SD	8.35	11.35	9.29	8.91	8.83	10.16	11.59
Arithmetic	Mean	12.48	13.83	13.70	14.83	14.15	12.83	11.33
	SD	3.26	3.41	3.59	3.62	3.14	2.74	2.93
Symbol Search	Mean	36.78	39.71	33.67	34.68	30.45	26.45	22.58
	SD	7.61	10.24	8.45	7.52	7.62	7.50	6.78
Visual Puzzles	Mean	17.78	17.63	16.86	16.48	14.18	11.78	9.58
	SD	3.83	3.99	4.46	4.64	4.31	4.06	3.19
Information	Mean	9.65	12.24	13.56	14.78	15.35	12.68	11.48
	SD	3.79	4.77	4.87	4.39	4.82	4.42	5.69
Coding	Mean	66.65	87.63	62.70	64.05	58.58	49.60	42.33
	SD	15.88	9.91	16.77	17.51	14.29	14.32	15.72
Letter-number sequencing	Mean	20.25	20.83	20.77	20.45	19.68	18.33	n/a
	SD	2.48	2.55	2.84	2.42	2.07	3.40	
Figure Weights	Mean	17.08	16.85	14.65	14.95	13.28	9.93	n/a
	SD	5.65	4.53	4.08	4.59	4.49	4.17	
Comprehension	Mean	21.18	21.85	24.98	24.48	25.65	22.10	19.65
	SD	5.23	5.82	4.25	5.72	4.73	5.67	5.96
Cancellation	Mean	39.48	39.71	37.02	37.08	34.62	28.90	n/a
	SD	9.97	8.93	11.29	10.30	9.65	9.09	
Picture Completion	Mean	14.10	14.59	13.91	15.18	12.90	11.15	9.45
	SD	3.28	3.09	3.05	3.15	3.37	3.89	3.52

**Table 3. Raw scores with their standard score and z-score equivalents. Age group 1 (16 to 20 years)**

Z	Score: Scaled	Raw Score													PC	
		BD	Sim	DS	MR	Voc	Ari	SS	VP	Inf	CD	LNS	FW	Com		Canc
-3.00	1	18.83	11.08	15.02	7.47	-	2.62	13.95	6.29	-	19.01	12.81	.13	5.49	9.57	4.26
-2.67	2	22.10	12.36	16.6	8.75	2.27	3.78	16.47	7.56	-	24.17	13.63	2.0	7.22	12.87	5.35
-2.33	3	25.46	13.66	18.21	10.07	5.11	4.89	19.05	8.86	0.82	29.65	14.48	3.92	9.0	16.25	6.46
-2.00	4	28.72	14.94	19.78	11.34	7.86	5.96	21.56	10.12	2.07	34.89	15.29	5.78	10.72	19.54	7.54
-1.67	5	31.99	16.21	21.36	12.62	10.67	7.04	24.08	11.39	3.33	40.14	16.11	7.65	12.45	22.84	8.63
-1.33	6	35.35	17.51	22.97	13.94	13.46	8.15	26.66	12.69	4.61	45.44	16.96	9.57	14.23	26.22	9.74
-1.00	7	38.61	18.78	24.54	15.21	16.21	9.22	29.17	13.95	5.86	50.77	17.77	11.43	15.95	29.51	10.82
-0.67	8	41.88	20.06	26.12	16.49	18.97	10.3	31.69	15.22	7.12	56.02	18.59	13.3	17.68	32.81	11.91
-0.33	9	45.24	21.36	27.73	17.81	21.81	11.41	34.27	16.52	8.4	61.41	19.44	15.22	19.46	36.19	13.02
0	10	48.50	22.63	29.30	19.08	24.56	12.48	36.78	17.78	9.65	66.65	20.25	17.08	21.18	39.48	14.10
+0.33	11	51.76	23.9	30.87	20.35	27.31	13.55	39.29	19.04	10.9	71.89	21.06	18.94	22.9	42.77	15.18
+0.67	12	55.12	25.2	32.48	21.67	30.15	14.66	41.87	20.34	12.18	77.28	21.91	20.86	24.68	46.15	16.29
+1.00	13	58.39	26.48	34.06	22.95	32.91	15.74	44.39	21.61	13.44	82.53	22.73	22.73	26.41	49.45	17.38
+1.33	14	61.65	27.75	35.63	24.22	35.66	16.81	46.9	22.87	14.69	87.77	23.54	24.59	28.13	52.74	18.46
+1.67	15	65.01	29.05	37.24	25.54	38.5	17.92	49.48	24.17	15.97	93.16	24.39	24.59	28.13	56.12	19.57
+2.00	16	68.28	30.33	38.82	26.82	41.26	19.0	52.0	25.44	17.23	98.41	25.21	28.38	31.64	59.42	20.66
+2.33	17	-	31.6	40.39	-	44.01	20.07	54.51	26.7	18.48	103.65	26.02	-	33.36	62.73	21.74
+2.67	18	-	32.9	42.00	-	46.85	21.18	57.09	-	19.76	109.04	26.87	-	35.14	66.12	22.85
+3.00	19	-	34.18	43.58	-	49.61	22.26	59.61	-	21.02	114.29	27.69	-	36.87	69.39	23.94

Note: In scoring an individual select the raw score closest to that of the individual. Numbers in *italics* are scores above the maximum possible rawtotal score for a subtest.

**Table 4. Raw scores with their standard score and z-score equivalents. Age group 2 (21 to 30 years, inclusive)**

Z	Score: Scaled	Raw Score													PC	
		BD	Sim	DS	MR	Voc	Ari	SS	VP	Inf	CD	LNS	FW	Com		Canc
-3.00	1	22.05	6.02	17.19	7.54	-	3.6	8.99	5.66	-	57.9	13.18	3.26	4.39	12.92	5.32
-2.67	2	25.12	7.95	18.57	8.78	-	4.73	12.37	6.98	-	61.17	14.02	4.75	6.31	15.87	6.34
-2.33	3	28.29	9.93	19.99	10.07	3.75	5.88	15.85	8.33	-	64.54	14.89	6.30	8.29	18.90	7.39
-2.00	4	31.36	11.86	21.37	11.32	7.5	7.01	19.23	9.65	2.7	67.81	15.73	7.79	10.21	21.85	8.41
-1.67	5	34.43	13.79	22.75	12.57	11.25	8.14	22.61	10.97	4.27	71.08	16.57	9.28	12.13	24.80	9.43
-1.33	6	37.60	15.77	24.17	13.85	15.10	9.29	26.09	12.32	5.90	74.45	17.44	10.83	14.11	27.83	10.48
-1.00	7	40.67	17.7	25.55	15.1	18.85	10.42	29.47	13.64	7.47	77.72	18.28	12.32	16.03	30.78	11.5
-0.67	8	43.74	19.63	26.93	16.35	22.60	11.55	32.85	14.96	9.04	80.99	19.12	13.81	17.95	33.73	12.52
-0.33	9	46.91	21.61	28.35	17.63	26.46	12.70	36.33	16.31	10.67	84.36	19.99	15.36	19.93	36.76	13.57
0	10	49.98	23.54	29.73	18.88	30.2	13.83	39.71	17.63	12.24	87.63	20.83	16.85	21.85	39.71	14.59
+0.33	11	53.05	25.47	31.10	20.13	33.95	14.95	43.09	18.95	13.81	90.9	21.67	18.35	23.77	42.66	15.61
+0.67	12	56.22	27.45	32.53	21.41	37.81	16.11	46.57	20.30	15.44	94.27	22.54	19.89	25.75	45.69	16.66
+1.00	13	59.29	29.38	33.91	22.66	41.55	17.24	49.95	21.62	17.01	97.54	23.38	21.38	27.67	48.69	17.69
+1.33	14	62.36	31.31	35.29	23.91	45.30	18.37	53.33	22.94	18.58	100.81	24.22	22.87	29.59	51.59	18.70
+1.67	15	65.53	33.29	36.71	25.19	49.15	19.52	56.81	24.29	20.21	104.18	25.09	24.42	31.57	54.62	19.75
+2.00	16	68.6	35.22	38.09	26.44	52.9	20.65	60.19	25.61	21.78	107.45	25.93	25.91	33.49	57.57	20.77
+2.33	17	-	37.15	39.47	-	56.65	21.78	-	26.93	23.35	110.72	26.77	27.40	35.41	60.52	21.79
+2.67	18	-	-	40.89	-	60.5	22.93	-	-	24.98	114.09	27.64	-	37.39	63.55	22.84
+3.00	19	-	-	42.27	-	-	-	-	-	26.55	117.36	28.48	-	-	66.5	23.86

Note: In scoring an individual select the raw score closest to that of the individual. Numbers in *italics* are scores above the maximum possible raw total score for a subtest.



**Table 5. Raw scores with their standard score and z-score equivalents. Age group 3 (31 to 40 years, inclusive)**

Z	Score: Scaled	Raw Score										Com	Canc	PC		
		BD	Sim	DS	MR	Voc	Ari	SS	VP	Inf	CD				LNS	FW
-3.00	1	20.59	13.06	13.78	2.16	8.06	2.93	8.32	3.48	-	12.39	12.25	2.41	12.23	3.15	4.76
-2.67	2	23.55	14.42	15.60	3.79	11.13	4.11	11.11	4.95	.56	17.92	13.19	3.76	13.63	6.88	5.77
-2.33	3	26.59	15.82	17.47	5.47	14.28	5.34	13.98	6.47	2.21	23.63	14.15	5.14	15.08	10.71	6.80
-2.00	4	29.55	17.18	19.28	7.1	17.35	6.52	16.77	7.94	3.82	29.16	15.09	6.49	16.48	14.44	7.81
-1.67	5	32.51	18.54	21.10	8.73	20.42	7.70	19.56	9.41	5.43	34.69	16.03	7.84	17.88	18.17	8.82
-1.33	6	35.55	19.94	22.97	10.41	23.57	8.93	22.43	10.93	7.08	40.40	16.99	9.22	19.33	22.00	9.85
-1.00	7	38.51	21.3	24.78	12.04	26.64	10.11	25.22	12.4	8.69	45.93	17.93	10.57	20.73	25.73	10.86
-0.67	8	41.47	22.66	26.6	13.67	29.71	11.29	28.01	13.87	10.3	51.46	18.87	11.92	22.13	29.46	11.87
-0.33	9	44.51	24.06	28.47	15.35	32.86	12.52	30.88	15.39	11.95	57.17	19.83	13.30	23.58	33.29	12.90
0	10	47.47	25.42	30.28	16.98	35.93	13.7	33.67	16.86	13.56	62.7	20.77	14.65	24.98	37.02	13.91
+0.33	11	50.43	26.78	32.10	18.61	39.0	14.88	36.46	18.33	15.17	68.23	21.71	16.0	26.38	40.75	14.92
+0.67	12	53.47	28.18	33.97	20.29	42.15	16.11	39.33	19.85	16.82	73.94	22.67	17.38	27.83	44.58	15.95
+1.00	13	56.43	29.54	35.78	21.92	45.22	17.29	42.12	21.32	18.43	79.47	23.61	18.73	29.23	48.31	16.96
+1.33	14	59.39	30.90	37.60	23.55	48.29	18.47	44.91	22.79	20.04	85.00	24.55	20.08	30.63	52.04	17.97
+1.67	15	62.43	32.30	39.47	25.23	51.44	19.70	47.67	24.3	21.69	90.71	25.51	21.46	32.08	55.87	19.00
+2.00	16	65.39	33.66	41.28	26.86	54.51	20.88	50.57	25.78	23.3	96.24	26.45	22.81	33.48	59.6	20.01
+2.33	17	68.35	35.02	43.10	-	57.58	22.06	53.36	27.25	24.91	101.77	27.39	24.16	34.88	63.33	21.02
+2.67	18	-	36.42	44.97	-	-	-	56.23	-	26.56	107.48	28.35	25.54	36.32	67.16	22.05
+3.00	19	-	-	46.78	-	-	-	59.02	-	-	113.01	29.29	26.89	-	70.89	23.06

Note: In scoring an individual select the raw score closest to that of the individual. Numbers in *italics* are scores above the maximum possible raw total score for a subtest.

**Table 6. Raw scores with their standard score and z-score equivalents. Age group 4 (41 to 50 years, inclusive)**

Z	Score: Scaled	Raw Score										Com	Canc	PC		
		BD	Sim	DS	MR	Voc	Ari	SS	VP	Inf	CD				LNS	FW
-3.00	1	19.06	13.85	15.32	5.87	11.47	3.97	12.12	2.56	1.61	11.52	13.19	1.18	7.32	6.18	5.73
-2.67	2	22.14	15.17	16.97	7.23	14.41	5.16	14.60	4.09	3.06	17.30	13.99	2.69	9.21	9.58	6.77
-2.33	3	25.31	16.54	18.68	8.62	17.44	6.4	17.16	5.67	4.55	23.25	14.81	4.26	11.15	13.08	7.84
-2.00	4	28.39	17.86	20.33	9.98	20.38	7.59	19.64	7.2	6.0	29.03	15.61	5.77	13.04	16.48	8.88
-1.67	5	31.47	19.18	21.98	11.34	23.32	8.78	22.12	8.73	7.45	34.81	16.41	7.28	14.93	19.88	9.92
-1.33	6	34.64	20.55	23.69	12.73	26.35	10.02	24.68	10.31	8.94	40.76	17.23	8.85	16.87	23.38	10.99
-1.00	7	37.72	21.87	25.34	14.09	29.29	11.21	27.16	11.84	10.39	46.54	18.03	10.36	18.76	26.78	12.03
-0.67	8	40.80	23.19	26.99	15.45	32.23	12.40	29.64	13.37	11.84	52.32	18.83	11.87	20.65	30.18	13.07
-0.33	9	43.97	24.56	28.7	16.84	35.26	13.64	32.2	14.95	13.33	58.27	19.65	13.44	22.59	33.68	14.14
0	10	47.05	25.88	30.35	18.2	38.2	14.83	34.68	16.48	14.78	64.05	20.45	14.95	24.48	37.08	15.18
+0.33	11	50.13	27.20	32.00	19.56	41.14	16.02	37.16	18.01	16.23	69.83	21.25	16.46	26.37	40.48	16.22
+0.67	12	53.30	28.57	33.71	20.95	44.17	17.26	39.72	19.59	17.72	75.78	22.07	18.03	28.31	43.98	17.29
+1.00	13	56.38	29.89	35.36	22.31	47.11	18.45	42.2	21.12	19.17	81.56	22.87	19.54	30.2	47.38	18.33
+1.33	14	59.46	31.21	37.01	23.67	50.05	19.64	44.68	22.65	20.62	87.34	23.67	21.05	32.09	50.78	19.37
+1.67	15	62.63	32.58	38.72	25.06	53.0	20.88	47.24	24.23	22.11	93.29	24.49	22.62	34.03	54.28	20.44
+2.00	16	65.71	33.9	40.37	26.42	56.02	22.07	49.72	25.76	23.56	99.07	25.29	24.13	35.92	57.68	21.48
+2.33	17	68.79	35.22	42.02	-	58.96	-	52.2	27.29	25.0	104.85	26.09	25.64	37.81	61.08	22.52
+2.67	18	-	36.59	43.73	-	-	-	54.76	-	26.50	110.80	26.91	27.21	-	64.58	23.59
+3.00	19	-	-	45.38	-	-	-	57.24	-	-	116.58	27.71	-	-	67.98	24.63

Note: In scoring an individual select the raw score closest to that of the individual. Numbers in *italics* are scores above the maximum possible raw total score for a subtest.

**Table 7. Raw scores with their standard score and z-score equivalents. Age group 5 (51 to 60 years, inclusive)**

Z	Score: Scaled	Raw Score														
		BD	Sim	DS	MR	Voc	Ari	SS	VP	Inf	CD	LNS	FW	Com	Canc	PC
-3.00	1	12.90	9.87	15.50	-0.17	13.31	4.73	7.59	1.25	0.89	15.71	13.47	-0.19	11.46	5.67	2.79
-2.67	2	15.81	11.59	16.94	1.44	16.22	5.77	10.10	2.67	2.48	20.43	14.15	1.29	13.02	8.85	3.90
-2.33	3	18.80	13.37	18.41	3.11	19.23	6.83	12.70	4.14	4.12	25.38	14.86	2.82	14.63	12.14	5.05
-2.00	4	21.71	15.09	19.85	4.73	22.14	7.87	15.21	5.56	5.71	30.00	15.54	4.30	16.19	15.32	6.16
-1.67	5	24.62	16.81	21.29	6.34	25.05	8.91	17.72	6.98	7.30	34.72	16.22	5.78	17.75	18.50	7.27
-1.33	6	27.61	18.59	22.76	8.01	28.06	9.97	20.32	8.45	8.94	39.57	16.93	7.31	19.36	21.79	8.42
-1.00	7	30.52	20.31	24.20	9.63	30.97	11.01	22.83	9.87	10.53	44.29	17.61	8.79	20.92	24.97	9.53
-0.67	8	33.43	22.03	25.64	11.25	33.88	12.05	25.34	11.29	12.12	49.01	18.29	10.27	22.48	28.15	10.64
-0.33	9	36.42	23.81	27.11	12.91	36.89	13.11	27.93	12.76	13.76	53.86	19.00	11.80	24.09	31.44	11.79
0	10	39.33	25.53	28.55	14.53	39.80	14.15	30.45	14.18	15.35	58.58	19.68	13.28	25.65	34.62	12.90
+0.33	11	42.24	27.25	29.99	16.15	42.71	15.19	32.96	15.60	16.94	63.30	20.36	14.76	27.21	37.80	14.01
+0.67	12	45.23	29.03	31.46	17.81	45.72	16.25	35.56	17.07	18.58	68.15	21.07	16.29	28.82	41.09	15.16
+1.00	13	48.14	30.75	32.90	19.43	48.63	17.29	38.07	18.49	20.17	72.87	21.75	17.77	30.38	44.27	16.27
+1.33	14	51.05	32.47	34.34	21.05	51.54	18.33	40.58	19.91	21.76	77.59	22.43	19.25	31.94	47.45	17.38
+1.67	15	54.04	34.24	35.81	22.71	54.55	19.39	43.18	21.38	23.40	82.44	23.14	20.78	33.55	50.74	18.53
+2.00	16	56.95	35.97	37.25	24.33	57.46	20.43	45.69	22.80	24.99	87.16	23.82	22.26	35.11	53.92	19.64
+2.33	17	59.86	37.69	38.69	25.95	-	21.47	48.20	24.22	26.58	91.88	24.50	23.74	36.67	57.10	20.75
+2.67	18	62.85	-	40.16	27.61	-	22.53	50.80	25.69	-	96.73	25.21	25.27	-	57.10	21.90
+3.00	19	65.76	-	41.60	-	-	53.31	53.31	27.11	-	101.45	25.89	26.75	-	63.57	23.01

Note: In scoring an individual select the raw score closest to that of the individual. Numbers in *italics* are scores above the maximum possible raw total score for a subtest.

**Table 8. Raw scores with their standard score and z-score equivalents. Age group 6 (61 to 70 years, inclusive)**

Z	Score: Scaled	Raw Score														
		BD	Sim	DS	MR	Voc	Ari	SS	VP	Inf	CD	LNS	FW	Com	Canc	PC
-3.00	1	7.45	11.41	13.06	-1.59	4.80	4.61	3.95	-0.40	-0.58	6.64	8.13	-2.58	5.09	1.63	-0.52
-2.67	2	10.42	12.73	14.52	-0.06	8.15	5.51	6.43	0.94	0.88	11.37	9.25	-1.20	6.96	3.82	0.76
-2.33	3	13.48	14.08	16.03	1.51	11.61	6.45	8.97	2.32	2.38	16.23	10.41	0.21	8.89	7.72	2.09
-2.00	4	16.45	15.40	17.49	3.04	14.96	7.35	11.45	3.66	3.84	20.96	11.53	1.59	10.76	10.72	3.37
-1.67	5	19.42	16.72	18.95	4.57	18.31	8.25	13.93	5.00	5.30	25.69	12.65	2.97	12.63	13.72	4.65
-1.33	6	22.48	18.07	20.46	6.14	21.77	9.19	16.48	6.38	6.80	30.27	13.81	4.38	14.56	16.81	5.98
-1.00	7	25.45	19.39	21.92	7.67	25.12	10.09	18.95	7.72	8.26	35.28	14.93	5.76	16.43	19.81	7.26
-0.67	8	28.42	20.71	23.38	9.20	28.47	10.99	21.43	9.06	9.72	40.01	16.05	7.14	18.30	22.81	8.54
-0.33	9	31.48	22.06	24.89	10.77	31.93	11.93	23.98	10.44	11.22	44.87	17.21	8.55	20.23	25.90	9.87
0	10	34.45	23.38	26.35	12.30	35.28	12.83	26.45	11.78	12.68	49.60	18.33	9.93	22.10	28.90	11.15
+0.33	11	37.42	24.70	27.81	13.83	38.63	13.73	28.93	13.13	14.14	54.33	19.45	11.31	23.97	31.90	12.43
+0.67	12	40.48	26.05	29.32	15.40	42.09	14.67	31.48	14.50	15.64	59.19	20.61	12.72	25.90	35.00	13.76
+1.00	13	43.45	27.37	30.78	16.93	45.44	15.57	33.95	15.84	17.10	63.92	21.73	14.10	27.77	37.99	15.04
+1.33	14	46.42	28.69	32.24	18.46	48.79	16.47	36.43	17.18	18.86	68.65	22.85	15.48	29.64	40.99	16.32
+1.67	15	49.48	30.04	33.75	20.03	52.25	17.41	38.98	18.56	20.06	73.51	24.01	16.89	31.57	44.08	17.65
+2.00	16	52.45	31.36	35.21	21.56	55.60	18.31	41.45	19.90	21.52	78.24	25.13	18.27	33.44	47.08	18.93
+2.33	17	55.42	32.68	36.67	23.09	58.95	19.21	43.93	21.24	22.98	82.97	26.25	19.65	35.31	50.08	20.21
+2.67	18	58.48	34.03	38.18	24.66	-	20.15	46.48	22.62	24.48	87.83	27.41	21.06	37.24	53.17	21.54
+3.00	19	61.45	35.35	39.64	26.19	-	21.05	48.95	23.96	25.94	92.56	28.53	22.44	-	56.17	22.82

Note: In scoring an individual select the raw score closest to that of the individual. Numbers in *italics* are scores above the maximum possible raw total score for a subtest.

**Table 9.** Subtest raw scores with their standard score and z-score equivalents. Age group 7 (71 years or more)

Z	Score:	Raw Score														
		BD	Sim	DS	MR	Voc	Ari	SS	VP	Inf	CD	LNS	FW	Com	Canc	PC
-3.00	1	-1.0	8.95	10.26	-2.49	-1.57	2.54	2.24	0.01	-5.59	-4.83	N/A	N/A	1.77	N/A	-1.11
-2.67	2	2.48	10.44	11.72	-1.01	2.25	3.51	4.48	1.06	-3.71	0.36	N/A	N/A	3.74	N/A	0.05
-2.33	3	6.07	11.97	13.23	0.51	6.20	4.50	6.78	2.15	-1.78	5.70	N/A	N/A	5.76	N/A	1.25
-2.00	4	9.55	13.45	14.69	1.99	10.02	5.47	9.02	3.20	0.10	10.89	N/A	N/A	7.73	N/A	2.41
-1.67	5	13.03	14.94	16.15	3.47	13.84	6.44	11.26	4.25	1.98	16.08	N/A	N/A	9.70	N/A	3.57
-1.33	6	16.62	16.47	17.66	4.99	17.79	7.43	13.56	5.34	3.91	21.42	N/A	N/A	11.72	N/A	4.77
-1.00	7	20.10	17.95	19.12	6.47	21.61	8.40	15.80	6.39	5.79	26.61	N/A	N/A	13.69	N/A	5.93
-0.67	8	23.58	19.44	20.58	7.95	25.43	9.37	18.04	7.44	7.67	31.80	N/A	N/A	15.66	N/A	7.09
-0.33	9	27.17	20.97	22.09	9.47	29.38	10.36	20.34	8.53	9.60	37.14	N/A	N/A	17.68	N/A	8.29
0	10	30.65	22.45	23.55	10.95	33.20	11.33	22.58	9.58	11.48	42.33	N/A	N/A	19.65	N/A	9.45
+0.33	11	34.13	23.94	25.02	12.43	37.02	12.30	24.82	10.63	13.36	47.52	N/A	N/A	21.62	N/A	10.61
+0.67	12	37.72	25.47	26.52	13.95	40.97	13.29	27.12	11.72	15.29	52.86	N/A	N/A	23.64	N/A	11.81
+1.00	13	41.20	26.95	27.98	15.43	44.79	14.26	29.36	12.77	17.17	58.05	N/A	N/A	25.61	N/A	12.97
+1.33	14	44.68	28.44	29.44	16.91	48.61	15.23	31.60	13.82	19.05	63.24	N/A	N/A	27.57	N/A	14.13
+1.67	15	48.27	29.97	30.95	18.43	52.56	16.22	33.90	14.91	20.98	68.58	N/A	N/A	29.60	N/A	15.33
+2.00	16	51.75	31.45	32.41	19.91	56.38	17.19	36.14	15.96	22.86	73.77	N/A	N/A	31.57	N/A	16.49
+2.33	17	55.23	32.94	33.87	21.39	60.20	18.16	38.38	17.01	24.74	78.96	N/A	N/A	33.54	N/A	17.65
+2.67	18	58.82	34.47	35.38	22.91	64.15	19.15	40.68	18.10	26.67	84.30	N/A	N/A	35.56	N/A	18.85
+3.00	19	62.30	35.95	36.84	24.39	67.97	20.12	42.92	19.15	-	89.49	N/A	N/A	37.53	N/A	20.01

Notes. Letter-Number Sequencing, Figure Weights and Cancellation subtests are not administered to examinees ≥ 70 years. In scoring an individual select the raw score closest to that of the individual. Numbers in *italics* are scores above the maximum possible raw total score for a subtest.

## DISCUSSION

The purpose of this paper was to produce normative data for Māori on the WAIS-IV, in a format accessible and applicable for clinical practice. The impetus for this work arose from an extensive body of literature that indicates current neuropsychological tests and practices, and their accompanying normative data introduces bias when applied to individuals who diverge from non-western cultures and may therefore lead to spurious diagnoses (Brickman et al., 2006; Fernández et al., 2008; Rivera Mindt et al., 2010; Uzzell et al., 2007). This is of particular concern for Māori here in Aotearoa who are culturally dissimilar to the western worldview in which the WAIS-IV was developed (Dudley et al., 2019; Ogden & McFarlane, 1997; Ogden et al., 2003).

Variables such as acculturation, education, and socio-economic (SES) status are complex, culturally influenced indicators and have been consistently identified as contributing to variance in neuropsychological performance (Ardilla, 2007; Arentoft, et al., 2015; Coffey et al., 2005; Manly et al., 1998; Kennepohl et al., 2004; Razani et al., 2007; Walker, Batchelor & Sores, 2010).

The WAIS-IV has been shown to draw heavily on the Western educational experience of the individual (Walker et al., 2010). This is also problematic for Māori who have historically been subjected to a western education system that has marginalised their culture, and ignored Māori pedagogy (Bishop et al., 2009; Gordon, 2018; Durie, 1998; Walker 2016) leading to poor education outcomes. (Statistics New Zealand, 2013; Bishop, 2009; Ministry of Education, 2013). Therefore, the discrimination and disadvantage that Māori have experienced in the education system continues to disadvantage Māori in the practice of neuropsychology.

In general, cross-cultural neuropsychologists dispute the notion that the variation in performance seen in some cross-cultural studies is due to genetic or biological differences and argue that differences are a product of political or social determinants as demonstrated in several studies where socio-economic factors have been controlled for (Evans et al., 2000; Ibanez-Casas, et al., 2016; Noble et al., 2007). Disappointingly, the significance of SES appears to continue to be downplayed or overlooked in the field of neuropsychology as indicated by a review of 1277 neuropsychology research journals between 2016 and 2019 that found only 13% of the articles provided the socio-economic status of the sample populations (Medina et al., 2021). In clinical practice it is critical for neuropsychologists to conduct a comprehensive evaluation that includes enquiry into the various factors

of the socioeconomic background of the person being assessed, to identify those variables that potentially impact performance.

This oversight extends to the WAIS-IV manual which does not include stratification of factors such as an individual's adaptation to the dominant western culture or their socio-economic status, thereby creating possible bias when administered to Māori. In a previous publication using the same sample, the degree to which a person identified with Māori culture, their education, and income levels were found to account for some of the variance in their performance on the Test of Premorbid Functioning (ToPF) (Dudley et al., 2017).

In response to the consistent requests from Māori whānau for a Māori-friendly environment when undergoing a neuropsychological assessment (Dudley et al., 2014; Dudley & Faleafa, 2016; Ogden & McFarlane, 1997; Ogden et al., 2003; Shepherd & Leathem, 1999), the current study employed culturally appropriate protocols of engagement and rapport-building that honour a Māori worldview. For example, offering the participants a choice of where to hold the interviews, offering the opportunity to have karakia, whanaungatanga and the sharing of kai all led to a sense of feeling included for the participant where otherwise that may have experienced feelings of exclusion and alienation. Incorporating these cultural practices can have a profound effect on the assessment procedure for the participant and a positive influence on their performance. Neuropsychologists in Aotearoa are ethically bound to promote an assessment environment in which the person is motivated to perform well. The hui process (Lacey et al., 2011), and the Meihana Model (Pitama et al., 2017) are two approaches that promote cultural awareness and provide for cultural safety practices and would be well suited for the administration of a neuropsychological assessment with Māori.

A strength of the present data includes the administration of the full WAIS-IV to a large, stratified sample, which was representative of the main iwi from across Aotearoa. To our knowledge, it is the first large scale effort to produce normative data for Māori on any version of the WAIS.

## Conclusion

This manuscript presents normative data tables for clinicians to use in scoring the WAIS-IV when administered to Māori. Our hope is that clinicians in Aotearoa will access this data to make fair comparison of an individual's performance against a selection of test scores derived from the administration of the WAIS-IV to a sample that is representative of the Māori population.

## References

- Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Ardilla, A. (2007). The impact of culture on neuropsychological test performance. In B. P. Uzzell, M. Ponton, and A. Ardilla, Eds. *International handbook of cross-cultural Neuropsychology*, 44. NJ: Lawrence Erlbaum.
- Arentoft, A., Byrd, D., Monzones, J., Coulehan, K., Fuentes, A., Rosario, A., ... & Rivera Mindt, M. (2015). Socioeconomic status and neuropsychological functioning: associations in an ethnically diverse HIV+ cohort. *The Clinical Neuropsychologist*, 29(2), 232-254.
- Bishop, R., Berryman, M., Cavanagh, T., & Teddy, L. (2009). Te kotahitanga: Addressing educational disparities facing Māori students in New



- Zealand. *Teaching and Teacher Education*, 25(5), 734-742.
- Brickman, A. M., Cabo, R., & Manly, J. J. (2006). Ethical issues in cross-cultural neuropsychology. *Applied Neuropsychology*, 13(2), 91-100.
- Coffey DM, Marmol L, Schock L, Adams W. (2005). The influence of acculturation on the Wisconsin Card Sorting Test by Mexican Americans. *Archives of Clinical Neuropsychology*; 20(6):795-803.
- Dudley, M. D., Barker-Collo, S. L., Wilson, D. L., & Garrett, N. K. (2019). Factors Associated with Māori Performance on the WAIS-IV. *Archives of Clinical Neuropsychology*, 34(7), 1203-1216.
- Dudley, M., Faleafa, M., & Anderson, E. (2016). Neuropsychological Assessment in New Zealand. In J. Rucklidge, J. Feather, M. Waitoki & N. Robertson. (Eds.), *Practice issues for Clinical and Applied Psychologists in New Zealand*. New Zealand Psychological Society: Wellington.
- Dudley M, Scott K, Barker-Collo S. (2017) Is the test of premorbid functioning a valid measure for Māori in New Zealand? *New Zealand Journal of Psychology*, 46(3):72-79.
- Dudley, M., Wilson, D., & Barker-Collo, S. (2014). Cultural invisibility: Māori people with traumatic brain injury and their experiences of neuropsychological assessments. *New Zealand Journal of Psychology*, 43(3).
- Durie, Mason (1998). *Whaiora, Maori Health Development* (2nd ed.). OxfordUniversity Press.
- Evans, J. D., Walden-Miller, W. S., Byrd, D. A., & Heaton, R. K. (2000). Cross-cultural applications of the Halstead-Reitan batteries. In E. Fletcher-Janzen, T.L. Strickland, & C.R. Reynolds, (Eds.), *Handbook of Cross-Cultural Neuropsychology* (pp. 287-303). Springer, Boston, MA.
- Feigin, V. L., Barker-Collo, S. L. (2007) Normative neuropsychological data: Do we need them in New Zealand? *New Zealand Family Physician*, 34 (6), 441-443.
- Feigin, V., K. Carter, et al. (2006) "Ethnic disparities incidence of stroke subtypes: Auckland Regional Community Study" *Neurology: The Lancet*, 5(1):130-139.
- Fernández, A.L., Abe, J. (2018). Bias in cross-cultural neuropsychological testing: problems and possible solutions. *Culture and Brain* 6, 1-35. <https://doi.org/10.1007/s40167-017-0050-2>
- Franzen, S., Papma, J. M., van den Berg, E., & Nielsen, T. R. (2021). Cross-cultural neuropsychological assessment in the European Union: a Delphi expert study. *Archives of Clinical Neuropsychology: the official journal of the National Academy of Neuropsychologists*, 36(5), 815-830.
- Gordon, J. (2018). Critical pedagogy in a Māori-medium setting. *Te Kaharoa*, 11(1).
- Haitana, T., Pitama, S., & Rucklidge, J. J. (2010). Cultural biases in the Peabody Picture Vocabulary Test-III: Testing tamariki in a New Zealand sample. *New Zealand Journal of Psychology (Online)*, 39(3).
- Huria, T., Palmer, S. C., Pitama, S., Beckert, L., Lacey, C., Ewen, S., & Smith, L. T. (2019). Consolidated criteria for strengthening reporting of health research involving indigenous peoples: the CONSIDER statement. *BMC medical research methodology*, 19(1), 1-9.
- Ibanez-Casas, I., Daugherty, J., Mejia, A., Leonard, B., Kuwabara, H., Hidalgo Ruzzante, N., ... & Puente, A. (2016). C-73 Relevance of Subjective Socioeconomic Status Measures for Cross-Cultural Neuropsychological Performance: The EMBRACED Project. *Archives of Clinical Neuropsychology*, 31(6), 667-667.
- Kennepohl, S., Shore, D., Nabors, N., & Hanks, R. A. (2004). African American acculturation and neuropsychological test performance following traumatic brain injury, *Journal of the International Neuropsychological Society*, 10(4):566-77.
- Lacey, C., Huria, T., Beckert, L., Gilles, M., & Pitama, S. (2011). The Hui Process: a framework to enhance the doctor-patient relationship with Māori. *New Zealand Medical Journal*, 124(1347), 72-8.
- Manly, J. J., Miller, S. W., Heaton, R. K., Byrd, D., Reilly, J. U. D. Y., Velasquez, R. J., ... & HIV Neurobehavioral Research Center (HNRC) Group. (1998). The effect of African American acculturation on neuropsychological test performance in normal and HIV-positive individuals. *Journal of the International Neuropsychological Society*, 4(3), 291-302.
- Manly, J. J. (2005). Advantages and disadvantages of separate norms for African Americans. *The Clinical Neuropsychologist*, 19(2), 270-275.
- Manly J. J. (2008). Critical issues in cultural neuropsychology: profit from diversity. *Neuropsychology Review*, 18(3), 179-183. <https://doi.org/10.1007/s11065-008-9068-8>
- McNaughton, H., M. Weatherall, et al. (2002) "The comparability of community outcomes for European and non-European survivors of stroke in New Zealand" *New Zealand Medical Journal*, 115(1149):98-100.
- Medina, L. D., Torres, S., Gioia, A., Lopez, A. O., Wang, J., & Cirino, P. T. (2021). Reporting of demographic variables in neuropsychological research: An update of O'Bryant et al.'s trends in the current literature. *Journal of the International Neuropsychological Society*, 27(5), 497-507.
- Ministry of Education. (2013). *Profile and trends: New Zealand's tertiary education sector 2012*. Wellington: Ministry of Education. [https://www.educationcounts.govt.nz/publications/tertiary\\_education/annual/2531/profile-and-trends-2013](https://www.educationcounts.govt.nz/publications/tertiary_education/annual/2531/profile-and-trends-2013)
- Ministry of Health. Cabinet Office circular CO (19) 5 Te Tiriti o Waitangi/Treaty of Waitangi Guidance 22 October 2019. <https://dpmc.govt.nz/publications/co-19-5-te-tiriti-o-waitangi-treaty-waitangi-guidance>
- Ministry of Health, NZ (2018) *Māori Health*, Ministry of Health, Wellington, <https://www.health.govt.nz/our-work/populations/maori-health>
- Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental science*, 10(4), 464-480.
- Ogden, J. A., Cooper, E., & Dudley, M. (2003). Adapting neuropsychological assessments for minority groups: A study comparing White and Māori New Zealanders. *Brain Impairment*, 4(2), 122-134.
- Ogden, J. A., & McFarlane-Nathan, G. (1997). Cultural bias in the neuropsychological assessment of young

- Māori men. *New Zealand Journal of Psychology* 26(2), 2-12.
- Pedraza, O. (2020). *Introduction to clinical cultural neuroscience*. Oxford University Press.
- Pedraza, O., & Mungas, D. (2008). Measurement in cross-cultural neuropsychology. *Neuropsychology review*, 18(3), 184-193.
- Pitama, S. G., Bennett, S. T., Waitoki, W., Haitana, T. N., Valentine, H., Pahiina, J., ... & McLachlan, A. (2017). A proposed hauora Māori clinical guide for psychologists: Using the hui process and Meihana model in clinical assessment and formulation.
- Razani, J., Burciaga, J., Madore, M., & Wong, J. (2007). Effects of acculturation on tests of attention and information processing in an ethnically diverse group. *Archives of Clinical Neuropsychology*, 22(3), 333-341.
- Rivera Mindt, M., Byrd, D., Saez, P., & Manly, J. (2010). Increasing culturally competent neuropsychological services for ethnic minority populations: a call to action. *The Clinical Neuropsychologist*, 24(3), 429-453. <https://doi.org/10.1080/13854040903058960>
- Ross-McAlpine, K. S., Leathem, J. M., & Flett, R. A. (2018). A survey of psychologists administering cognitive and neuropsychological assessments with New Zealand children. *New Zealand Journal of Psychology*, 47(1).
- Shepherd, I., & Leathem, J. (1999). Factors affecting performance in cross-cultural neuropsychology: From a New Zealand bicultural perspective. *Journal of the International Neuropsychological Society*. 5(1), 83-84.
- Statistics New Zealand (2013). <https://stats.govt.nz/tereocensus/previous-censuses/2013-census/>
- Tan, Y. W., Burgess, G. H., & Green, R. J. (2021). The effects of acculturation on neuropsychological test performance: A systematic literature review. *The Clinical Neuropsychologist*, 35(3), 541-571.
- Uzzell, B. P., Ponton, M., & Ardila, A. (Eds.). (2013). *International handbook of cross-cultural neuropsychology*. Psychology Press.
- van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, 54, 119-135.
- Walker, A Batchelor, J., & Shores, EA. (2010). Effects of cultural background on WAIS-III and WMS-III performances after moderate-severe traumatic brain injury. *Australian Psychologist*, 45(2), 112-122.
- Walker, R. (2016). Reclaiming Māori education. *Decolonisation in Aotearoa: Education, Research, and Practice*, 19-38.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV)*. The Psychological Corporation.
- Zawaly, K., Moyes, S. A., Wood, P. C., Cheung, G., Rolleston, A., Buetow, S., ... & Kerse, N. (2019). Diagnostic accuracy of a global cognitive screen for Māori and non-Māori octogenarians. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 5, 542-552.

**Corresponding Author**

Makarena Dudley

Email: [m.dudley@auckland.ac.nz](mailto:m.dudley@auckland.ac.nz)

School of Psychology, University of Auckland

# Stakeholders' Perceptions of a New Zealand Youth Mentoring Programme Assisting High-Achieving, Underprivileged Students to Attend University

Hana Turner-Adams and Melinda Webber

Faculty of Education and Social Work, University of Auckland, New Zealand

This mixed-methods research study utilised interviews and online surveys to investigate a New Zealand-based youth mentoring programme that assists academically talented students to attend university. The study explored stakeholders' perceptions of the programme's benefits and challenges through data collected from current students and alumni (n = 144), mentors (n = 137), and financial partners (n = 49). Most participants expressed satisfaction with the programme. A noteworthy strength was the programme's three-pronged approach of financial support with university costs, a mentor for each student, and paid work experience. Challenges with the programme included mentors' and financial partners' lack of cultural knowledge about the Māori and Pacific Island students they supported. Mentors also appeared ill-equipped to deal with communication and relationship breakdowns with their mentees.

**Keywords:** *Youth mentoring, first-in-family, cultural awareness, higher education, mentor training*

## INTRODUCTION

Attaining a university qualification provides graduates with multiple benefits. University graduates earn more money, secure higher-level jobs, and are less likely to be unemployed than those without qualifications (Ministry of Education, 2021). Graduates also have a greater sense of purpose and life satisfaction (Universities New Zealand, 2017). In New Zealand, fewer Māori, Pacific Island, and students from low socioeconomic schools attend university than Pākehā/NZ European students and those from higher socioeconomic backgrounds (Education Counts, 2021). Therefore, it is essential to identify and implement strategies that support underprivileged students attending and succeeding at university.

Youth mentoring is a strategy that has been introduced into some schools and communities in New Zealand to help address educational inequities. Mentoring programmes are often offered to students who attend schools in low socioeconomic communities or those from disadvantaged backgrounds. Mentoring intends to provide students with support and access to opportunities to improve their educational outcomes and career prospects.

Formal youth mentoring began in New Zealand in the early 1990s (Farruggia et al., 2011) and is defined as "a long-term relationship between a younger, less experienced individual and an older, more experienced individual who provides guidance in a particular domain" (Evans & Ave, 2000, p. 41). Mentoring benefits young people in a range of ways. A synthesis of 55 studies on youth mentoring found that mentored young people benefitted significantly in the domains of emotional, psychological, problematic/high-risk behaviour, social competence, academic/educational, career progression, and employment (DuBois, Holloway, et al., 2002).

## Successful Mentoring Programmes

A mentoring programme's success depends on a range of interacting features. Prior research has shown that recruiting mentors from caring or helping professions for their relationship-building experience is more effective than recruiting from other occupations (DuBois, Holloway, et al., 2002). Clear expectations about the level of contact between mentors and mentees, planned activities, parental involvement, and careful monitoring and evaluation also feature in successful programmes (DuBois, Holloway, et al., 2002; Dubois, Neville, et al., 2002).

Jacobi (1991) identified five components of effective mentoring: First is an achievement focus where the mentor supports the mentee to succeed in education or work. Second, the mentor provides career guidance and emotional support. Third, the relationship is reciprocal and beneficial to both the mentor and mentee. Fourth, the mentoring relationship involves regular interaction, communication, and collaboration, and finally, mentors need to have more experience, influence, and achievement within their field than their mentees.

In a New Zealand study, Dutton et al. (2018) also recognised the critical components of a quality mentoring relationship. Like Jacobi (1991), quality mentoring focused on improvement for the mentee in areas such as academic performance, self-esteem or self-efficacy. The mentor and mentee were also invested in the relationship and worked collaboratively. Other features included a mutually respectful bond between the mentor and the mentee; a shared purpose and goals; and a relationship that improved over time through learning from and about each other.

Programmes that incorporate mentoring alongside other interventions appear to be more successful at

meeting their goals than sole mentoring programmes (Jolliffe & Farrington, 2007; LoSciuto et al., 1996). For example, a study by LoSciuto et al. (1996) combined mentoring of low-income students with a life skills course and regular interactions with older people in residential care. They found that students who met regularly with their mentors were absent less often than students involved at an average or marginal level. Mentored students were also more optimistic about their future, schooling, and participation in community service. These findings from successful mentoring programmes highlight the importance of the mentoring pair committing to and nurturing a mutually beneficial and respectful relationship that leads to positive changes for the mentee.

### **Problems with Mentoring**

Despite the benefits of youth mentoring programmes, the overall effect sizes are small (DuBois, Holloway, et al., 2002). In addition, some research has found that "the mentoring relationship can be detrimental to the mentor, the mentee or both" (Long, 1997, p. 115). In particular, ending the mentoring relationship early or relationships that last less than one year reduces the effectiveness and adversely affects students' well-being (Farruggia et al., 2011; Grossman et al., 2012).

### **Matching of Mentors and Mentees**

Research has examined whether matching demographic characteristics led to better mentoring relationships. DuBois, Holloway, et al. (2002) did not find that matching mentoring pairs on attributes or interests were significant moderators of effect size. However, other studies have found that matching led to better relationships and more positive outcomes for mentored youth (Ensher & Murphy, 1997; Raposa et al., 2018). A shared dislike of activities also predicted a longer youth-mentor relationship than shared interests or when interests differed. Race and ethnicity matches had a lower risk of relationship termination due to the mentor moving away or losing interest but a higher risk of termination due to conflict (Raposa et al., 2018). An ethnic mismatch between mentor and mentee may be problematic if the mentee has cultural mistrust towards people from ethnic groups with whom there is a history of colonisation, racism or discrimination. Allowing mentored youth and their parents to share their preferences for mentors before matching may alleviate these risks (Sánchez et al., 2021; Sánchez et al., 2013).

### **Culturally-focused Mentoring**

There is limited research on culturally responsive youth mentoring, which is surprising, given that New Zealand mentoring programmes recruit significant numbers of Māori and Pacific Island youth (Farruggia, Bullen, Davidson, et al., 2011). Additionally, more than 50% of mentoring programmes, even those developed primarily for Māori, do not feature whānau (family) involvement, which suggests that programme organisers may not recognise the importance of whānau, hapū, and iwi connections (Farruggia, Bullen, Solomon, et al., 2011).

In a recent New Zealand study, Ualesi (2021) explored culturally responsive, sustaining, and safe mentoring practices. She identified seven ingredients needed to

support Māori and Pacific Island youth effectively. These include: (1) "A culturally safe space; (2) positive social identities; (3) covenant relationships; (4) culture of self-determination; (5) culture of honour; (6) culturally transformative relationships; and (7) sacred space" (p.170). Indigenous knowledge systems are often missing from Western mentoring models, but Ualesi's study highlighted the necessity of cultural centrality for mentoring programmes to benefit Māori and Pacific Island youth.

### **The Current Study**

The current study is focused on a New Zealand youth mentoring programme that targets academically talented students from financially disadvantaged backgrounds. The programme aims to support students to achieve their academic potential through tertiary education, so they can positively influence and benefit their communities. Limited research exists about programmes that combine mentoring with other interventions and support students through the transition from school to tertiary education. Research in the New Zealand context is also scarce. Therefore, the current study aimed to investigate the efficacy of a multi-component youth mentoring programme in New Zealand from the perspective of its key stakeholders.

Students in the programme receive four years of individual mentoring from their last year of high school to the third year of their degree, university tuition fees, and paid work experience. With a focus on academically talented students, applicants need to achieve high grades in the National Certificate of Educational Achievement (see New Zealand Qualifications Authority, n.d.). The household income threshold for the scholarship and entry to the mentoring programme is NZ\$75,000 for families with one dependant and NZ\$90,000 for larger families. Priority is given to students who are the first in their families to attend university. Once students are in the programme, they must pass all university courses and maintain a B grade or higher. From the second and third years of university, students are expected to maintain a B+ average or better. In addition, they must submit all assessments and attend all examinations required for each course.

Mentors volunteer to join the programme. They complete induction training before starting mentoring and commit to supporting a student for a minimum of two years. Mentors and mentees are encouraged to meet at least once a month in addition to regular contact by phone, email, or text message. Participation in programme-organised mentoring activities and networking events is encouraged. The mentoring pair completes a quarterly progress and feedback report, and mentors have access to a support team in the programme organisation if problems arise.

Financial and work experience partners (partners) are representatives of organisations that provide financial support to students in the programme by contributing to their university course costs. Although the scholarship primarily covers course fees, the partner may allow some of the funds to be allocated to other costs, such as textbooks, transport, accommodation, and IT resources. Partner organisations also provide students with a

minimum of 4-5 weeks of paid work experience per year and employment-based mentoring and support.

Through interviews and surveys with each of the programme's stakeholders, multiple perspectives and experiences allowed cross-validation of important themes across the programme. Two research questions were central to the study: (1) What are the benefits/strengths of the mentoring programme? (2) What are the challenges associated with the mentoring programme, and how could these be addressed?

**METHOD**

**Study design**

We used a pre-experimental, post-test-only design to explore stakeholder perceptions of a multi-faceted youth-mentoring programme. Although the design lacks a control or comparison group, it is common in studies that assess the efficacy of an intervention, like that of the mentoring programme in the current study (Cervera et al., 2020).

**Materials and Procedure**

The research was conducted in two stages: First, in-depth interviews with current students, alumni, mentors, and partners. Each interview lasted between 30 and 50 minutes and took place face-to-face or by telephone. For stage two, an invitation to participate in an online survey was sent to all students, alumni, mentors, and partners.

The surveys comprised rating scales and open-ended questions specific to each type of stakeholder (student, alumni, mentors, and partners). For the rating scales, participants indicated agreement or disagreement with different prompts. Example prompts for students about their mentor were, 'They are a good match for me' and 'They are never too busy to meet up with me'. Examples of open-ended interview and survey questions included:

- What do you think are the benefits of having a mentor? (Students and alumni)
- Overall, what do you think are the highlights of mentoring students? (Mentors)
- Are there any issues you have experienced with your students? (Partners).

**Participants**

The interview participants (N = 15) were purposively sampled and included alumni (n = 3) who had completed the programme in the previous 2-4 years, current students (n = 4), mentors (n = 5), and partners (n = 3). There were 330 participants for the survey, including alumni (n = 56), students (n = 88), mentors (n = 137), and partners (n = 49).

Demographic data were collected about students and alumni, but not for mentors and partners. The ethnicity and gender breakdown of students and alumni are presented in **Error! Reference source not found..**

**Table 1.** Gender and Ethnicity of Students and Alumni

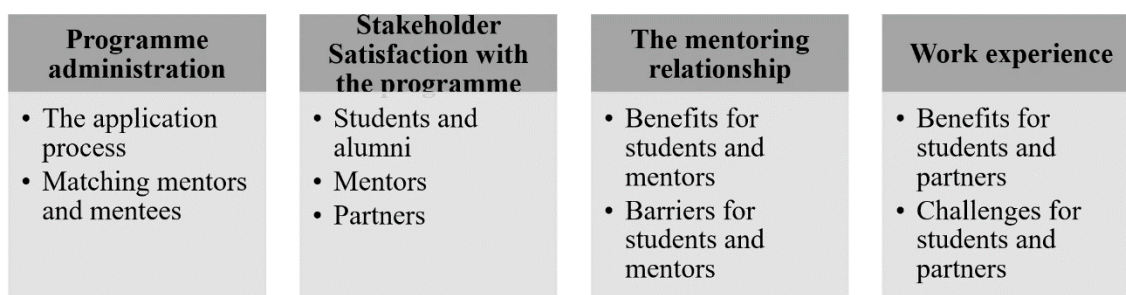
	Students (n = 88)	Alumni (n = 56)
<b>Gender</b>		
Female	68	35
Male	20	21
<b>Ethnicity</b>		
Māori	34	11
Pākehā	14	18
Pacific Islands	34	14
Indian (incl. Fijian Indian)	0	6
Asian	3	4
Other	3	3

**Data Analysis**

The interview and survey data were analysed thematically using the steps outlined in Braun and Clarke's (2006) approach for analysing qualitative data, which involves "identifying, analysing, and reporting patterns (themes) within data" (Braun & Clarke, 2006, p. 79). Thematic analysis is a flexible method of analysing data that suits research related to people's experiences, perceptions or viewpoints. The six phases of thematic analysis are "(1) Familiarisation with the data; (2) Generating initial codes; (3) Searching for themes; (4) Reviewing the themes; (5) Defining and naming themes; (6) Producing the report" (Braun & Clarke, 2006, p. 87).

The final themes and sub-themes are displayed in **Error! Reference source not found..** The first theme, 'Programme administration', focused on the application and interview process and how mentors were matched to mentees. The second theme, 'Programme satisfaction', described stakeholders' perceptions of the overall programme. The third theme, 'The mentoring relationship', included mentors and students' perceptions of the benefits and challenges with mentoring. Finally, the fourth theme, 'Work experience', focused on student and partner-identified benefits and challenges with the work experience aspect of the programme.

**Figure 1.** Study Themes and Sub-themes





## RESULTS

### Programme Organisation and Administration

*The application process:* Students learned about the mentoring programme through school newsletters or friends and predominantly drove the application process themselves rather than being approached to apply. Students reported they had received sufficient information about the programme and understood the commitment and expectations. One student said, "They told us what to wear to the interview and to thoroughly prepare... They gave us a call after each interview and asked for feedback and said the wait would be a month or so...yeah, they kept us informed" (Interview).

Some students found the application process time-consuming and the interview daunting. However, the interviewers were friendly, the feedback was positive, and the process taught students valuable job application skills. One student said, "[It] was one of the hardest interviews I've ever had to do ... [but it] prepares you for future interviews..." (Interview).

The 'speed-dating' format of the entry interviews limited what some students were willing to share about themselves in front of others. One student said,

*There was a bunch of us in one room, and it kind of felt like we were listening to one another; it made us compare what each other said. If it was one-to-one, we could talk more and concentrate on what we were saying with no eavesdropping* (Interview).

The partners were affirming of the interviews and agreed that a stringent process was necessary to ensure students took it seriously and that the recipients they selected were deserving. One partner explained:

*It is a hard process because there is a financial component... We expect students to present to us, to do a 10-minute presentation... But let's not forget that kids do this – it's their bread and butter – they already have to do this at school. But if they want \$20, they have to work for it, and they have to get up and [have the] ability to speak. I think it's a very robust process. It's a point of difference, and it's great. It gives us high-quality students* (Interview).

Although students' awards and grades were of interest during the selection process, one respondent said the letter written by students which explained their reasons for wanting to be in the programme was most valuable:

*They write a letter, "why I am applying", and someone writes them a reference. Those two are what we are really interested in... The letters often cover what's in the forms anyway. [A well-written letter] gives us the same information in their own voice* (Interview).

*Matching mentors with mentees:* Overall, mentors were satisfied with the matching process and perceived it was effective. One mentor said, "My buddy and I get on so well (due to the matching), and the support from [the programme] is amazing" (Survey). Mentors who had experienced the "speed dating" interviews supported the format. One mentor said,

*[The] speed-dating process to match up was good. Students had a few minutes to talk about themselves, [and] it gives the students an element of choice... Some people didn't get picked, but that's the nature of the game. It's fascinating to see people's backgrounds* (Interview).

Another mentor who had not participated in speed dating and had just been given students to mentor said, "As with anything, it's tough. Some people won't click, but I don't know if anything will improve the odds" (Interview).

### Stakeholder Satisfaction with the Programme

*Alumni and students:* Alumni (96%) and students (78%) agreed that the mentoring programme had met or exceeded their expectations, with only a few who said it had failed. The youth mentoring programme in the current study was unlike other scholarships available to students because of the work experience and mentoring provided. Several students had not previously held jobs. They perceived that the work experience aspect of the mentoring programme advantaged them over other job applicants, gave them an increased sense of self-direction, and enabled them to develop skills in the workforce. However, as students had not previously received mentoring, they were initially unsure of its potential benefits.

Overall, most current students (61%) and alumni (41%) identified that all three aspects of the programme (i.e., work experience, mentor, and financial assistance) were most valuable. However, the remaining alumni chose work experience (24%) as most valuable, which was slightly higher than financial aid (22%). Mentoring was of least value to alumni (10%). For the remaining current students, financial support was the main attraction (17%) as it reduced the need for a student loan. Their mentor (14%) rated more highly than the work experience (8%). Students new to the programme were not likely to have spent as much time doing work experience as alumni. Without the benefit of time to reflect, they may not have fully realised its value.

*Mentors:* Mentors were primarily part of the programme to give back to the community, as some had received mentoring themselves, and others wanted to help those less fortunate in life. A quarter of mentors surveyed (25%) reported a positive experience with the programme, and 16% found the organisation supportive and helpful. Several mentors described the programme as inspiring and hopeful and referred to its high success rate, which they perceived was due to the financial, pastoral, and employment support. One mentor said, "...All three points of the triangle are essential to get the students through what they need. The success rate is very high because of this, particularly compared to other scholarship schemes" (Interview).

Mentors liked that the programme targeted talented, driven students and offered a 'hand up', not a 'hand-out' to help them reach their potential. One mentor said, "I appreciated what they were doing in terms of taking people who didn't have university role models. It wasn't just throwing money at people; they have mentors that can assist, that can provide that extra assistance" (Interview).

*Partners:* Partners also had a high level of satisfaction with the programme (80% rated the programme as 4 or 5, where five is *very well* and zero is *very poor*). They identified several reasons for being involved: (1) to enable talented students to realise their leadership potential; (2) to give back to the community; (3) to provide a hand-up, not a hand-out; (4) the wrap-around approach of the programme; and (5) the opportunity for their company to gain perspectives from a younger generation. More than just corporate social responsibility, partners viewed the programme as holistically assisting promising young people to increase their chances of success. The 'hand up' philosophy also resonated strongly with partners who relished the opportunity to support talented students to reach their potential.

### **The Mentoring Relationship**

*Student-identified benefits:* The most frequently mentioned benefit for students was mentors' advice about personal matters, careers, schoolwork, and work experience. One student said, "It's good having someone who I can ask for help when I get stuck on assignments..." (Survey). Another student appreciated emotional support. S/he said, "I consider her more of a friend. I feel comfortable enough to discuss anything with her, even personal problems, and she has great advice..." (Survey). Some students had underestimated the value of the mentoring relationship at the beginning of the programme. However, by the end of the programme, many viewed the relationship positively. Students referred to their mentor as "a friend", "my second grandmother", "amazing", or "part of the family" (Interviews).

Students identified critical components in effective mentoring relationships. First, the student and mentor were both satisfied with the frequency and type of contact and support. Appropriate communication and support differed amongst the students. Some preferred face-to-face meetings or fun activities with their mentors, whereas others were happy with texts or phone calls. Equally, some students needed a significant level of support, whereas other more self-sufficient students required less. Second, the mentoring relationship involved academic/work advice and a personal connection. For example, mentors and students met each other's families. Finally, the mentoring relationship provided the student with connections, networks or expertise that directly helped the student. For example, a recruitment manager helped students with their curriculum vitae, and a lecturer advised them on appropriate university courses.

*Mentor-identified benefits:* Most mentors found mentoring highly rewarding. They enjoyed contributing to their student's life and felt pride and satisfaction as they saw them grow in confidence. One mentor appreciated "the satisfaction of seeing someone flourish... [I] went to the Māori presentation with [student name] and was blown away with how included they made me feel" (Survey). Student graduation was also a momentous occasion. For example, "Seeing a student graduate [was] a huge moment" (Survey).

Some mentors saw their students as friends, family members, or like their own children. One mentor said, "[I am] very proud of where my first girl has ended up, but she did it all on her own" (Survey). Students gave mentors

a new understanding of the younger generation, someone from a different culture, and an insight into the hardships some experienced.

### **Barriers to the Mentoring Relationship**

*Student-identified barriers:* Infrequent contact and communication were barriers to developing mentoring relationships for some students. One said, "I think we could keep in contact more" (Survey). Another student said, "We should try to find other ways to catch up, such as phone/video call, social media, etc." (Survey). Often, busy schedules prevented students and mentors from connecting. One student reported, "We find it really hard to meet up during the week as our timetables clash, or something comes up" (Survey). Some students were reticent to contact mentors who appeared very busy. One said, "I feel like I'm the one who always instigates when to meet up..." but she also added, "I understand that she is busy" (Survey). Other students' mentors were in a different town/city, which meant limited opportunities for face-to-face contact.

Personality clashes were another barrier that led to some students needing to change mentors. One student who had encountered problems with her mentor reported that she had not informed programme organisers about the issues because she did not want to appear ungrateful. As she had not experienced mentoring previously, she was unsure what to expect. Comparatively, mentors identified a significantly higher number of barriers with the mentoring relationship.

*Mentor-identified barriers:* The obstacles that mentors encountered as part of the mentoring relationship included cultural differences, expectations, frequency and form of contact, and managing serious issues.

*Cultural differences:* Several mentors reported cultural differences between them and their students. These appeared to stem, in some cases, from the mentor's limited cultural knowledge. One mentor suggested that a 'Polynesian' student should have boundaries with his parents. The mentor said,

*Some of the issues of expectations on students, in the case of my current students, comes from a Polynesian family who had enormous expectations placed on them. Perhaps some support to help the student put boundaries around those...not to take on too much [and] dealing with issues of fear and wanting to please the parents* (Interview).

This quote indicated that the mentor might not have fully understood parents' and children's roles and responsibilities within a Pacific Island or Māori family. Other mentors referred to family access to the money that students received. Some comments implied that money should be kept separately or withheld to prevent parents from accessing it. For example, one mentor said,

*Would be a better idea to be more controlling of the finances...The firm could pay [the programme], and [the programme] could put away savings for that child and then release 50% back to the child, which can then put it into the family coffers* (Survey).

In contrast, one mentor recognised that cultural differences could be a barrier in the mentoring relationship

and suggested recruitment of more Māori and Pasifika mentors:

*It would be great to find more PI [Pacific Islands] and Māori mentors because it's old-fashioned that it's the white middle-class people teaching the ethnic minorities. [The programme] needs to make sure they are advertising in their communities because it would be a big help (Interview).*

*Differences in expectations.* Mentors' expectations differed regarding the role they should have in a student's life, as some needed a lot of support, whereas others were more independent. Some mentors were unsure how often communication should occur and the best form of contact. Although two-thirds of mentors surveyed (27%) responded that they would continue to support the programme, the most reported barrier (47% of mentors) was the required level of commitment. Some mentors expressed guilt at the lack of time available to commit to their students or about juggling other obligations when their students needed them. One mentor said, "My current student is in the process of dropping out of the programme. I wonder what I could have done differently... Could I have taken a different approach?" (Survey). However, it appeared that mentees were a low priority for some mentors. One mentor said, "[The] only low point is just fitting it into your life. I'm feeling guilty because I haven't caught up with my student for her birthday, but I've been swamped with work" (Survey).

*Communication issues:* Some mentors reported problems with communicating or establishing a connection with their students. One mentor explained her student had been difficult to contact. "We had a few hitches in that department – my student never had any phone credit...and it can be quite hard to get hold of her, but I alerted [the programme] when I was worried..." (Survey). Other mentors wanted further training. For example, one said, "[learning] how to engage with the student via social media...the best apps to use to communicate (e.g. Snapchat)" as one way that could enhance communication between mentors and students.

*Serious issues:* When problems arose in the students' lives, mentors dealt with difficulties ranging from relationship breakups to mental illness. Some mentors experienced the same disappointment and hurt that they imagined a parent might feel. For example, "When something sh\*\*y goes on in their life, you feel it with them" (Survey).

### **Work Experience**

*Benefits of work experience for students and partners:* Current students reported that flexible work hours, a good team, and increased confidence were positive aspects of their work experience. However, although alumni were optimistic about the benefits of exposure to working life, only around 50% of those surveyed were positive about the connections they had made or the impact of work experience on their career decisions.

Interviewed students all reported excellent relationships with their partners. They appreciated the work experience and perceived it gave them an advantage over other job applicants. One student stated, "If I do my part well enough, then I could possibly continue working

with them" (Interview). Some students highlighted that their role aligned with their studies, and they viewed the work experience as "a connection and pathway into a potential career" (Interview).

Several students enjoyed experiencing something completely different. Work experience provided opportunities to "work somewhere you may not otherwise have worked, and thus getting to meet people you may not otherwise have met" (Interview). Although jobs were often low-skilled tasks such as packing, filing, and working on reception, their responsibilities increased as students became more familiar with the company.

Partners reported that their staff benefitted from the students being in the company. For example, some staff members had taken on internal coaching/management, which was valuable for new graduates who might not otherwise have had this opportunity. Like mentors, some partners reported that the experience of working with students had opened their eyes to the hardships that some people encountered in their everyday lives. Additionally, partners benefitted from having access to intelligent, driven young people, which enabled the organisation to keep in touch with a younger generation who brought new perspectives. One partner said, "We get fantastic skills, bright intellectuals, leadership skills, extracurricular interests, fresh ideas, and thinking" (Interview).

### **Barriers of work experience for students and partners.**

*Communication and staff awareness:* Poor communication and staff awareness was a barrier for 24% of students who responded that the contact at their work experience placement appeared to be unaware of the programme. One student said, "It would be nice if the main person of contact were fully aware of the requirements so that both parties don't slack off on certain commitments" (Survey). Another student said that her colleagues at the work experience placement had no idea why she was there, and the only task she did at work experience was driving a forklift. She said,

*Forklift driving was pretty exciting...but, unfortunately, this is all I have been doing... I haven't been given a chance to explore other aspects of the workplace...I assume they are not fully aware of my purpose there as I explain myself over and over again [as to] why I'm there... (Survey).*

Partners also commented about the importance of effective communication, staff being well informed about the students' role(s) in the company, and how the work experience aligned with other support provided to students. One partner said, "We needed to work hard at profiling them so that the rest of the company remember they are a part of the team" (Interview).

*Work availability:* Some students reported that staff changes at their work experience placements reduced work availability. One student, who had been at two different companies, reported that no one at either place appeared to want the responsibility for organising her work. She said,

*I have been with two workplaces...It was going well with my first placement until the boss, who signed up for [the programme] left, and the new*



*boss neglected me to the point I couldn't get any work at all... History has begun to repeat itself...The boss who signed me up [in the second workplace] has left, and I'm beginning to be neglected once again* (Survey).

Although the financial contribution students received from working was valued and appreciated, students reported that the work given was often unrelated to their field of study (e.g. packing crates when studying psychology and criminology). Students also said that work was not always available, and occasionally, there were shifts where there was nothing for them to do.

Partners reported that work availability was reliant on student communication, and some students did not recognise the importance of advising their work experience placement in advance about their availability. One partner said, "Communication has been a minor issue. We encourage our students to keep us in the loop regarding their work commitments with us" (Interview). Another said there was a "lack of urgency to organise holiday work, leaving it until late in the year to request preferred locations, making work difficult to find" (Interview). The comments from students and partners indicated that communication needed to improve to ensure the work experience process worked smoothly for all parties.

*Maintaining a work-study balance:* Some students reported difficulties managing their partner's expectations around work and study. Partners did not always seem to understand that they were full-time university students. For example, one student said, "Management should remember that this is work experience" (Survey). Furthermore, partners needed to ensure students had enough time to keep on top of their university studies and not expect them to work too many hours.

Partners were aware that some students had difficulty balancing work and study, and consequently, they spent too much time working. One partner said they needed to "[Keep] them on track with their own commitments" and ensure that students were not "over-committing themselves to work when they should be studying" (Interview). Some partners reported that students who were having problems at university avoided studying and instead increased their work hours. One partner said, "When students find their degree far more challenging than they anticipated, they find full-time work more appealing than their study or they over-commit to their work experience" (Interview).

*Financial issues:* Several of the work experience placements were in a corporate environment, and some students did not have business wear and could not afford to buy new clothes. Partners also commented about students' financial difficulties and noted that students needed assistance with transport (e.g. providing students with a bus card) or obtaining clothes to 'fit in'. One partner said, "...we have had our eyes opened to the situations that some of these students are coming from...They come from deprived homes; for example, struggling to get a wardrobe together" (Interview).

Some partners had similar views to mentors about students' finances, which demonstrated limited knowledge of or disagreement with the concept of income sharing, which is common in collectivist cultures. One

partner had "...cultural challenges around funding responsibilities" and expressed concern about a student who could not always afford to take the bus to work because their earnings were used to support the entire family (Survey).

## DISCUSSION

The purpose of this research was to investigate the perceptions of students, mentors and partners who were involved in a multi-component youth mentoring programme. The programme provided academically talented students with mentoring, paid work experience, and financial support to cover their university fees. This section discusses the main findings alongside the existing literature, the study limitations, and suggestions for further research. Finally, recommendations are made for stakeholders involved in youth mentoring programmes.

### **Cultural understanding and matching of mentor pairs**

Some mentors and partners demonstrated limited knowledge about their mentees' ethnic/cultural backgrounds. For example, they suggested that students' scholarship funds or earnings were being 'misused' when students shared the money they earned with their families. In collectivist cultures, it is common for income earned by individual family members to be pooled or shared. Contributing financially to the family is also how Pacific Islands children fulfil obligations to their parents (Benseman et al., 2006).

In another situation, a mentor referred to a 'Polynesian' student needing to set boundaries with their parents. An important Pacific Island (and Māori) value is to respect parents and elders. For a mentor to suggest that students set limits with their parents appears to disrespect cultural values that differ from their own (Fletcher et al., 2009). Both comments demonstrated the paternalistic attitudes of some mentors and partners towards students from non-dominant ethnic groups. Existing research supports matching mentors with mentees from the same race or ethnicity (Ensher & Murphy, 1997; Raposa et al., 2018), which may alleviate some cultural misunderstandings. However, most mentors are from white/middle-class backgrounds (Evans & Ave, 2000), which could delay matching students with a mentor from the same culture. Therefore, it would be beneficial for mentoring organisations to provide cultural competency training to help mentors better understand and support their mentees (Ptak et al., 1995). Additionally, as part of building a respectful and reciprocal mentoring relationship, mentors must learn about and know their mentees and vice versa (Dutton, 2018). It would appear that a mentor learning about a mentee's culture is crucial to understanding who they are.

*Problem resolution within the mentoring relationship:* When mentors and students in this study encountered problems in the mentoring relationship, they were not always aware of the resolution process. In one scenario, a student had personality clashes with her mentor but did not initially speak up. According to the Youth Mentoring Network (2016), an organisation created to provide a hub for mentors and mentoring providers in New Zealand, programmes need to assist mentor pairs with problems related to any aspect of mentoring. Although some

mentors and mentees had a contact person within the programme organisation, others were unsure who to contact. The New Zealand Youth Mentoring Network (2016) also recommends that programmes have a clear and accessible complaints resolution process. In the current study, it did not appear that mentors or students had access to or were aware of a complaints process which, if needed, may help reduce the risk of a mentoring relationship ending early.

*Frequency of contact:* The main barrier for students was the lack of regular contact with their mentors. Irregular contact with mentors made some students feel like they were a low priority. Rhodes et al., 1994, cited in Evans and Ave (2000), noted that students who had previous experience of natural mentoring relationships, such as those with older siblings or grandparents were often quicker to develop relationships with mentors than students with no prior mentor experiences. Other students who formed relationships had mentors who kept showing up and who persevered with the relationship, even when the mentee exhibited an apparent lack of interest. Regular communication and interaction are essential for a successful mentoring relationship (Jacobi, 1991). Consequently, mentoring programmes need to recruit mentors willing to meet regularly with and support their mentees.

### **Work experience**

A unique feature of the mentoring programme explored in this study was the inclusion of 4-5 weeks of paid work experience per year. The work experience allowed students to earn additional income and obtain work-related skills. As many employers expect graduates to be 'work-ready', paid work experience during studying increases employability and provides financial benefits (Evans, 2021). However, having a part-time job while completing full-time study has been shown to have adverse effects on students' academic outcomes (Callender, 2008). Although students in the current study appreciated the work experience, they were also conscious that their studies needed to be prioritised. Therefore, work experience within a mentoring programme needs to work alongside students' studies, contribute to increased knowledge of their subject area (where possible), and provide financial assistance.

### **Limitations of this Study and Suggestions for Further Research**

The research design in this study was a pre-experimental, post-test-only design. A limitation of this design is that it lacks a control or comparison group. Including a control group of unsuccessful applicants could have shown the differences in outcomes between mentored and unmentored students. Additionally, the data collection from stakeholders occurred once they had joined the programme. Adding a pre-test to the design may have emphasised further benefits of the programme. Future research could include a longitudinal control group study that tracks participants' perceptions before, during, and after the mentoring programme intervention. Tracking students from their last year of high school to the

end of their university degrees would demonstrate whether mentored students' perceptions and outcomes differed from their non-mentored peers.

### **Conclusion**

This study focused on stakeholders' perspectives of a multi-component youth mentoring programme. It highlighted a range of benefits and challenges for mentors, student mentees and partners involved in these types of programmes. Based on the findings of this research, several recommendations for youth mentoring programmes and practitioners were apparent. As the more experienced adult in the relationship, it is recommended that mentors take the lead in contacting their mentees and arranging regular meeting times. A minimum meeting of once a month is suggested, but research shows that more frequent interactions are associated with positive, enduring mentoring relationships (LoSciuto et al., 1996). Mentoring pairs are also encouraged to call, email, or text between the pre-arranged meeting times.

A common concern for mentors was a breakdown in communication when problems arose. A problem-solving process is needed that is easily accessible and that students and mentors can follow if problems occur within the mentoring relationship. Both parties should know the steps to take if they cannot contact their mentoring partner or if one partner wants to end the relationship or withdraw from the programme. Additionally, each mentoring pair needs a designated staff member within the programme organisation who makes regular contact, and access to a general inquiry line if their usual staff member is unavailable.

A recommendation for mentoring programme organisers is to train mentors in relationship-building, effective communication, cultural competency and cross-cultural understanding to help alleviate some of the issues raised about financial and familial expectations. Partners would also benefit from training and induction in these areas to better support the students working in their organisations.

A further recommendation for partners and mentors relates to work experience placements and the importance of not imposing unreasonable expectations or demands on young school leavers. If support with business wear is needed, mentors could advise about work clothes or help students access charities that supply clothing for low-income individuals entering the workforce. Partners could provide uniforms for students (where appropriate), or donate clothes for students to wear if they are in public-facing roles. Providing transport, parking, or assisting with public transport costs would also benefit students.

This research adds to the small number of studies that focus on multi-component youth mentoring programmes. The findings from the research highlighted that a wrap-around programme of mentoring, work experience, and financial aid, when implemented effectively, enables talented, high-achieving young people from underprivileged backgrounds to access university education.

## References

- Benseman, J., Coxon, E., Anderson, H., & Anae, M. (2006). Retaining non-traditional students: lessons learnt from Pasifika students in New Zealand. *Higher Education Research & Development, 25*(2), 147–162.
- Callender, C. (2008). The impact of term-time employment on higher education students' academic attainment and achievement. *Journal of Education Policy, 23*(4), 359–377.
- Cervera, N., Diago, P. D., Orcos, L., & Yáñez, D. F. (2020). The acquisition of computational thinking through mentoring: An exploratory study. *Education Sciences, 10*(8), 1-11.  
<https://doi.org/https://doi.org/10.3390/educsci10080202>
- DuBois, D. L., Holloway, B., Valentine, J., & Cooper, H. (2002). Effectiveness of mentoring programs for youth: A meta-analytic review. *American Journal of Community Psychology, 30*(2), 157–197.  
<https://doi.org/https://doi.org/10.1023/A:1014628810714>
- Dubois, D. L., Neville, H. A., Parra, G. R., & Pugh-Lilly, A. (2002). Testing a new model of mentoring. *New Directions for Youth Development, 2002*(93), 21–57.  
<https://doi.org/https://doi.org/10.1002/yd.23320029305>
- Dutton, H., Bullen, P., & Deane, K. (2018). Getting to the heart of it: understanding mentoring relationship quality from the perspective of program supervisors. *Mentoring & Tutoring: Partnership in Learning, 26*(4), 400-419.  
<https://doi.org/https://doi.org/10.1080/13611267.2018.1530132>
- Education Counts. (2021). *Tertiary Participation*.  
<https://www.educationcounts.govt.nz/statistics/tertiary-participation>
- Ensher, E. A., & Murphy, S. E. (1997). Effects of race, gender, perceived similarity, and contact on mentor relationships. *Journal of Vocational Behavior, 50*(3), 460–481.
- Evans, C. (2021). Exploiting students' part-time work to enhance learning, teaching and assessment. *Industry and Higher Education, 35*(1), 10–13.
- Evans, I. M., & Ave, K. T. (2000). Mentoring children and youth: Principles, issues, and policy implications for community programmes in New Zealand. *New Zealand Journal of Psychology, 29*(1), 41–49.
- Farruggia, S. P., Bullen, P., Davidson, J., Dunphy, A., Solomon, F., & Collins, E. (2011). The effectiveness of youth mentoring programmes in New Zealand. *New Zealand Journal of Psychology 40*(3), 52–70.
- Fletcher, J., Parkhill, F., Fa'aoi, A., & O'Regan, B. (2009). Pasifika students: Teachers and parents voice their perceptions of what provides supports and barriers to Pasifika students' achievement in literacy and learning. *Teaching and Teacher Education 25*(1), 24–33.
- Grossman, J. B., Chan, C. S., Schwartz, S. E., & Rhodes, J. E. (2012). The test of time in school-based mentoring: The role of relationship duration and re-matching on academic outcomes. *American Journal of Community Psychology, 49*(1), 43–54.
- Jacobi, M. (1991). Mentoring and undergraduate academic success: A literature review. *Review of Educational Research, 61*(4), 505–532.
- Jolliffe, D., & Farrington, D. P. (2007). *A rapid evidence assessment of the impact of mentoring on re-offending: A summary*. Home Office.
- Long, J. (1997). The dark side of mentoring. *The Australian Educational Researcher, 24*(2), 115–133.
- LoSciuto, L., Rajala, A. K., Townsend, T. N., & Taylor, A. S. (1996). An outcome evaluation of Across Ages: An intergenerational mentoring approach to drug prevention. *Journal of Adolescent Research, 11*(1), 116–129.
- Ministry of Education. (2021). *Education, income and earnings - with updates for 2020*.  
[https://www.educationcounts.govt.nz/\\_data/assets/pdf\\_file/0008/208268/Education-income-and-earnings-2020-Report.pdf](https://www.educationcounts.govt.nz/_data/assets/pdf_file/0008/208268/Education-income-and-earnings-2020-Report.pdf)
- New Zealand Qualifications Authority. (n.d.). *How NCEA Works*. New Zealand Government.  
<https://www.nzqa.govt.nz/ncea/understanding-ncea/how-ncea-works/>
- New Zealand Youth Mentoring Network. (2016). *Guide to effective and safe practice in youth mentoring* (2nd ed.). Author.
- Ptak, C. L., Cooper, J., & Brislin, R. (1995). Cross cultural training programs: Advice and insights from experienced trainers. *International Journal of Intercultural Relations, 19*(3), 425–453.
- Raposa, E. B., Ben-Eliyahu, A., Olsho, L. E., & Rhodes, J. (2018). Birds of a feather: Is matching based on shared interests and characteristics associated with longer youth mentoring relationships? *Journal of Community Psychology, 47*(2), 1–13.  
<https://doi.org/10.1002/jcop.22127>
- Sánchez, B., Anderson, A. J., Weiston-Serdan, T., & Catlett, B. S. (2021). *Anti-Racism education and training for adult mentors who work with BIPOC adolescents* (Vol. 36). SAGE Publications.
- Sánchez, B., Colón-Torres, Y., Feuer, R., Roundfield, K. E., & Berardi, L. (2013). Race, ethnicity and culture in mentoring relationships. In D. L. DuBois & M. J. Karcher (Eds.), *Handbook of youth mentoring* (Second ed., pp. 145-158). SAGE Publications.
- Ualesi, Y. M. (2021). *Culturally Responsive, Sustaining and Safe Youth Mentoring Practice in Aotearoa New Zealand - A Va Relational Approach* [Unpublished PhD Thesis, University of Auckland].
- Universities New Zealand. (2017). *A degree is a smart investment*. <https://www.universitiesnz.ac.nz/latest-news-and-publications/degree-smart-investment-0>

## Corresponding Author

Hana Turner-Adams  
Email: [h.turner@auckland.ac.nz](mailto:h.turner@auckland.ac.nz)

## Acknowledgements

Data for this study were collected and disseminated by The Perceptive Group, Auckland. [www.perceptive.co.nz](http://www.perceptive.co.nz)

## Trust and legitimacy: Policing among racial groups

Mariska Kappmeier<sup>1</sup> and Kathryn H. Fahey<sup>2</sup>

<sup>1</sup>National Centre for Peace and Conflict Studies, University of Otago

<sup>2</sup>Psychology Department, University of Otago

When it comes to trust in the police, ethnicity matters: members from minority groups trust the police less than majority group members. Without trust the police lack legitimacy and consequently groups' cooperation. While trust and legitimacy are closely related constructs, less is known if the ethnic-based differences in trust and legitimacy hinges upon different dimensions. This work addresses this question by utilising the multidimensional Intergroup Trust Model, which identifies the five dimensions of competence, integrity, compassion, compatibility, and security as a comprehensive set constituting trust. Three hundred and fifty participants were surveyed in three Boston neighbourhoods. Through mediation and regression analysis, the study found that for White participants legitimacy was predicted by integrity-based trust. Black participants perceived police as less legitimate, which was predicted by the lack of compassion-based trust. The indication of the findings for policing in the US and NZ are discussed.

**Keywords:** *Intergroup Trust Model, trust in the police, legitimacy*

### INTRODUCTION

Police play an integral role in protecting society and keeping communities safe. But what happens when the very communities police are there to protect, fail to perceive police as a legitimate institution? It is likely that, failing to find authority legitimate correlates with a lower sense of trust in police (Jackson & Gau, 2016), and a lower likelihood of reaching out to police when threatened or in danger. The result may be increased exposure to crime and violence (Moravcová, 2016; Panditharatne et al., 2018). Concerningly, this is the reality for many ethnic minority groups. Indeed, research across the USA (e.g. Burgason, 2017; Mummolo, 2018; Tyler, 2005), United Kingdom (e.g. Griffiths, 2018), Belgium (e.g. Van Craen & Skogan, 2015), and Finland (e.g. Kääriäinen & Niemi, 2014), as well as closer to home in Australia (e.g. Sargeant et al., 2014) and New Zealand (e.g., Kappmeier, Guenoun, & Campbell, 2019; Panditharatne et al., 2018; Quince, 2007; Te Whaiti & Roguski, 1998), consistently shows ethnic minorities trust the police less than the majority group.

Lower trust in the police by minority groups often stems from historical antagonism, between oppressive and prejudicial police and victimised minorities. Indeed, past and present experiences of brutality, harassment, and bias create perceptions of the police as racially and/or culturally discriminatory, procedurally prejudiced, and ultimately untrustworthy (Burgason, 2017; Schuck et al., 2008; Sivasubramaniam et al., 2008). In the present day, the Black Lives Matter protests emerging across the USA and the globe, highlighted the centuries of prejudicial treatment minority groups faced at the hands of police (see Weine et al., 2020). Further, the arising "Defund the Police" slogan highlighted the perception of an illegitimate and untrustworthy police force.

Researchers and media alike have focused much of their attention on the strained relationships between police and minority groups in the USA, however, a similar

pattern may be observed here in Aotearoa. The global BLM protests from 2020, including in Aotearoa, reflects that the strained relations between ethnic minorities and police is not an isolated phenomenon. Indeed, in Aotearoa, Māori communities are less likely than Pākehā communities to report that they trust the police (e.g. Panditharatne et al., 2018). Further, Te Whaiti and Roguski (1998) highlights the negative consequences of the police's bias and discrimination towards Māori communities on Māori trust.

In addition to consistently reporting lower trust in police, minority group members are also less likely to perceive the police as legitimate (Tyler, 2010, 2011). The legitimacy of the police is based on the social contract that the police hold but not misuse the state monopoly on violence. This provides the foundation of consent philosophy for policing adopted by many Western countries (Jackson et al., 2013). In most Western countries it is the police who are tasked with enforcing the law, and who are allowed to use violence if necessary to achieve this. But for this social contract to work it is of utter importance that the police are seen to be following the rules, are being a legitimate institution, and are indeed proving themselves to be trustworthy.

Given the importance of communities trusting the police, and given the ethnic-based difference in this trust as outlined above, this paper examines how minority and majority group members differ in their examination of police legitimacy.

### *Trust and legitimacy*

Jackson and Gau (2016) differentiate between trust and legitimacy: that trust is based on how the police fulfil the function they are tasked with (how competent, how well in line with moral values etc), while legitimacy refers to the perception that the police rightfully hold the power to fulfil their duty. Following this conceptualisation of

trust and legitimacy of the police, we theorise a causal relation between trust to legitimacy.

Furthermore, Jackson and Gau (2016) distinguish in their conceptual model of legitimacy between duty to obey, and institutional trust. Duty to obey relates to the feeling that it is your positive duty to obey police instructions. Institutional trust relates to the idea that the community believes police are appropriate, have the requisite properties to justify the possession of their power, and can be trusted to wield their power judiciously. Indeed, legitimacy is considered a combination of a collective agreement that police are duly authorised to possess the power they do, and the institutional trust that police can use this power in a manner that is just (Jackson & Gau, 2016). Legitimacy, as referred to here, is considered a consequence of how police treat people, and how police make decisions when they are exercising their legal authority (Tyler, 2011).

When authority is defined as legitimate, the “duty to obey” replaces personal morality (Kelman & Hamilton, 1989). That is, when citizens view police as a legitimate authority, they allow police to define social boundaries and appropriate behaviour in a social context. If police are considered legitimate, citizens will voluntarily comply with police orders, as citizens trust that the orders police give are fair and just. Without legitimacy, police are not viewed as moral, just, and proper in their use of power. Lower legitimacy would suggest communities do not trust police to respond professionally, efficiently, and fairly to their cries for help (Jackson & Gau, 2016).

Overall, given the differences in experiences with police for ethnic minorities versus majorities, both in Aotearoa and internationally, it is not surprising that minority group members tend to exhibit differing levels of trust in, and legitimacy towards police. This is problematic, as lower trust in police results in lower cooperation with police and a lower likelihood of reaching out to police when they are in danger, resulting in a greater exposure to crime (Moravcová, 2016; Panditharatne et al., 2018). Failing to perceive police as legitimate is not only harmful to minority groups, it is also harmful to police themselves. When police officers investigate crimes in the community, they need to be seen as legitimate: as a just, trustworthy entity, not an oppressive force. When perceived as a legitimate institution, police are more likely to receive important support and cooperation from the community (Sargeant et al., 2014; Tyler, 2016; Tyler & Jackson, 2014; Murphy et al., 2018). A link between trust and legitimacy is strongly implied: legitimacy requires trust, and this trust must be earned. However, due to the multi-dimensionality of trust, how minority versus majority groups develop trust in police may differ.

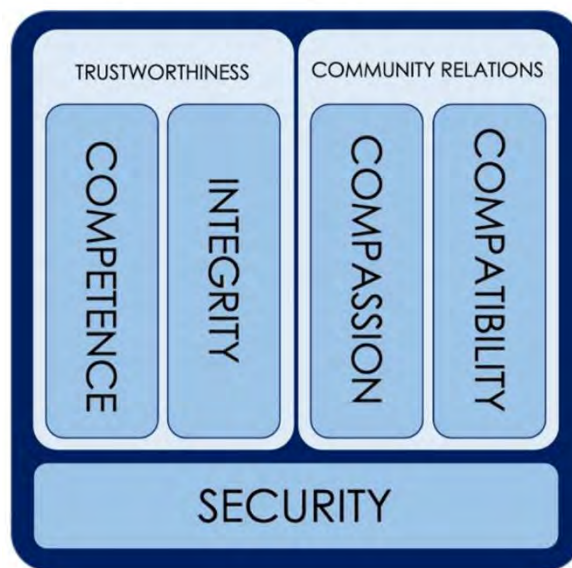
**Multi-dimensional approaches to trust**

Recent research has begun to understand trust in police as a multi-dimensional model (Balliet & Van Lange, 2013; Connelly et al., 2018; McEvily & Tortoriello, 2011; PytlikZillig & Kimbrough, 2016). Trust in the police can, for example, be based on the expectation that they have the skills, experience and reliability to keep communities safe: a competency-based trust. However, trust in the police can also be based on the expectation that police are honest and engage with the community

honourably: an integrity-based trust. Given this, the question examining the relationship between trust and legitimacy becomes not so much if trust predicts legitimacy of police, but rather which dimension of trust achieves this.

Kappmeier, Guenoun, & Fahey (2021) developed a five-dimensional trust model, the Intergroup Trust Model, which is particularly well suited to capture a more nuanced understanding of which type of trust predicts legitimacy (IGT-Model, see figure 1).

Figure 1. The Intergroup Trust Model



These five dimensions are broadly split into two categories: trustworthiness, and intergroup relations, with a third category of security-based trust relating to both of these. The trustworthiness category includes dimensions of competence and integrity. As mentioned above, the competence dimension of trust addresses perceptions of how effective or capable police are, whilst the integrity dimension addresses perceptions of whether the police are honest and guided by a moral code acceptable to one's own ingroup (Kappmeier, Guenoun & et al., 2019). The intergroup relations category includes dimensions of compatibility and compassion. Compassion addresses the perception of whether police care for your ingroup, whilst compatibility addresses the perception that one's group can relate to the police through shared language, culture, race, or experiences (Kappmeier, Guenoun & et al., 2019). Security is a dimension on its own, and addresses the perception that police will not harm my group physically or psychologically (Kappmeier, Guenoun & et al., 2019). Though the model is new, support for IGT-Model has been found across varied intergroup contexts (Kappmeier, Guenoun & et al., 2021).

**Race, legitimacy, and the Intergroup Trust Model**

While the Intergroup Trust Model (IGT-Model) was developed to assess trust in group settings (Kappmeier, 2016; Kappmeier, Guenoun & et al., 2021; Kappmeier, Venanzetti & Campbell, 2021), past research has also examined if ethnic minority groups based their trust in police on different trust dimension proposed by the IGT-



Model. Indeed, previous research using the IGT-Model for trust in the police found that while White Americans' trust in the police tend to be integrity-based: a stronger focus on honorability and morality of police. Conversely, Black American's reported a lack of compassion- and compatibility-based trust in the police: a stronger focus on the intergroup relational dimensions of trust (Kappmeier, Guenoun & et al., 2019). In other words, Whites trust in police tended to be based on how honest police were perceived to be, how moral police actions were perceived to be, and that police did not abuse their power. On the other hand, Blacks trust in police tended to be based on how compassionate police were in their interactions with Blacks, how concerned and attentive the police were for the Blacks needs, and how much police were perceived to be able to relate to Blacks (via background, language, traditions, values, beliefs, etc.). A similar pattern was found in Aotearoa: Māori's trust in the police was also shaped by an emphasis on compatibility-based trust (Kappmeier et al., 2019).

While previous work has used the IGT-Model to examine ethnic-based trust in the police, the link between the five dimensions of the IGT-Model and legitimacy has not yet been explored. Addressing this gap, the aim of this study is to examine ethnic-based differences in perceived legitimacy, specifically focusing on the distinct dimensions of trust recently identified through Intergroup Trust Model. To investigate this question, we conducted a community-based study in the context of the race relation in the United States of America, comparing Black Americans with White Americans.

**METHOD**

**Participants and procedure**

The community-based study took place in Boston USA, and data was collected in three neighbourhoods from August 2016 – December 2016. The three neighbourhoods were chosen for their similar socio-economic status, a known co-variate that influence trust in the police (e.g., Burgason, 2017), but differing racial demographic: one was predominantly Black, one predominantly White, and the third one had an approximately equal racial representation (see the supplementary material for a more detailed description on the three neighbourhoods, reasoning selections and study procedures). We deliberately chose to collect the data within community (vis-à-vis an online or student sample) to ensure that a) our participants are policed by the same department, particularly since policing approaches within the United States can differ broadly even between counties (President's Task Force on 21st Century Policing, 2015), b) to have a more homogenous policing experiences by recruiting from similar neighbourhoods, and c) avoid a recruitment bias often inherent to online participants sample. Finally, d) we also wanted to recruit participants from a broader breath than a student-based sample represented at the Higher Education.

A total of 372 respondents completed a pen and paper survey, at numerous locations in the neighbourhoods. To focus on minority-majority asymmetry only the responses of Black and White respondents were retained, since the two group memberships are the least ambiguous regarding their minority-majority status. The final sample included 252 respondents with 136 Black (61 female, 74 male, 1 unidentified) and 116 White participants (44 female, 72 male). The Black mean age was 36.1 (*SD* = 12.16; range 18 – 66) and the White mean age was 40.5 (*SD* = 15.9; range 19 – 87).

**Materials**

Participants responded regarding which racial group they most identified with.

**Trust measures:** Trust in the police was assessed through 19 items measuring the five dimensions of the IGT-Model (competence, 4 items; integrity, 6 items; compassion, 4 items; compatibility, 2 items; and security, 3 items). The items were displayed with opposite anchors on both sides. Participants indicated which side of the statement they strongly, somewhat, or slightly agreed with. For example, "We have nothing to fear from them" paired with "We have something to fear from them" (Security). This unusual form was chosen as it helps to lower multicollinearity, which has appeared in previous work (Kappmeier, Guenoun & et al., 2021). The alpha Cronbach was very good for competence ( $n = 4; \alpha = .70$ ); integrity ( $n = 6; \alpha = .80$ ); compassion ( $n = 4; \alpha = .86$ ) and security ( $n = 3; \alpha = .79$ ). Only for compatibility was it in the medium to good range ( $n = 2; \alpha = .67$ ), but given that the scale consisted of only two items, the Cronbach still indicates good reliability (Field, 2013).

**Legitimacy measure:** Legitimacy was assessed by: "Overall the police force is a legitimate institution and people should obey the decisions made by police officers" (Tyler, 2005), answered using a seven-point Likert scale (1 = strongly agree, 7 = strongly disagree).

**RESULTS**

**Preliminary analysis**

As a first step, descriptive statistics were analysed for legitimacy and dimensions of trust separately for each race. Black respondents reported lower legitimacy and lower trust in police across all five dimensions compared to White respondents (see table 1). Independent *t*-tests were conducted for legitimacy as well as all five dimensions within the IGT-Model. These findings show that Black respondents reported significantly lower trust in the police than White respondents on all five trust

**Table 1.** Descriptive statistics for minority and majority group members legitimacy and different dimensions of IGT-Model

	Blacks		Whites	
	$\bar{X}$	<i>sd</i>	$\bar{X}$	<i>sd</i>
Legitimacy	3.92	2.03	5.03	1.89
Competence	3.97	1.26	5.19	1.51
Integrity	3.48	1.22	4.90	1.40
Compassion	3.62	1.44	5.03	1.42
Compatibility	3.51	1.48	4.74	1.67
Security	3.46	1.51	5.02	1.67

**Table 2.** Independent t-tests for legitimacy of police and five trust dimensions across Black and White respondents.

	<i>t</i>	<i>df</i>	<i>LCI</i>	<i>UCI</i>	<i>d</i>
Legitimacy	4.15**	218	.584	1.641	0.55
Competence	-6.87**	239	-1.577	-.874	0.97
Integrity	-8.06**	219	-1.772	-1.076	1.15
Compassion	-8.53**	228	-1.993	-1.245	1.12
Compatibility	-8.53**	228	-1.993	-1.24	0.83
Security	-7.49**	231	-1.979	-1.154	1.03

\*\*p < .001; \*p < .05

dimensions, and significantly lower legitimacy of police (see table 2).

To examine ethnic-based differences in perceived legitimacy, we conducted two sets of analyses: After the preliminary analysis, we first completed a parallel mediation analysis with race as the predictor of legitimacy, mediated by the five trust dimensions, to assess the influence of race on legitimacy. In a second analysis we explored in more detail how the five dimensions of the IGT-Model predict legitimacy for Black and White participants respectively.

**Relationship between race and legitimacy, mediated by trust**

In order to understand the relationship between race and legitimacy, a multiple parallel mediation analysis, using ordinary least square was conducted. Race was the predictor, legitimacy was the outcome, and the five trust dimensions were mediators, modelling an indirect path from race to legitimacy. The analysis was conducted in SPSS, using the Haynes process tool 3.3, Model 4.

Notable in the parallel mediation analysis, the direct path from race to legitimacy was not significant ( $c' = -0.5, p = .59, CI [-.11, .15]$ ; see Figure 2 for details). This indicates that the race of participants did not predict how legitimate they perceived the police to be. However, a significant path from the race of the participants to all five trust dimensions, indicating that all five trust dimensions picked up on race-based differences in the participants (see Table 3 for details).

When examining the effect from each trust dimension to legitimacy, the results show that only the path from compassion-based trust was significant, suggesting that the five trust dimensions differ in their relevance for police legitimacy. While there is no direct effect of race on police legitimacy, there is an indirect effect via compassion-based trust. The more compassionate the police are perceived to be, the more legitimate they are perceived as, and this is especially true with respect to minority (over majority) group members.

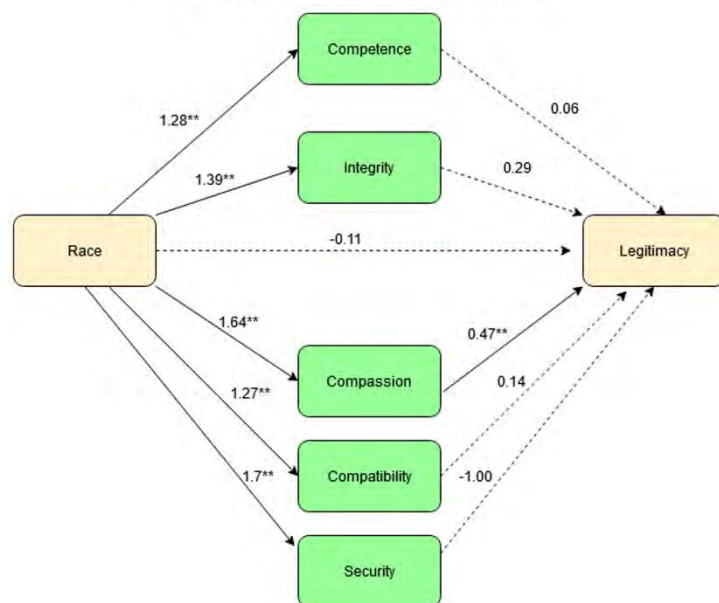
**Racial group members views on legitimacy via different dimensions of trust**

Given that race had a direct effect on all five trust dimensions, there is some suggestion that the dimension(s) of trust most influential with respect to legitimacy of police may differ for majority and minority group members. In order to test this, a multiple linear regression analysis was conducted, using legitimacy as the dependent variable (see table 4 for details).<sup>1</sup>

For Black participants, from the five dimensions explaining the variance of trust, only compassion was a significant predictor for legitimacy ( $\beta = 0.74, p = .002$ ). None of competence ( $\beta = -0.11, p = .598$ ), integrity ( $\beta = 0.13, p = .627$ ), compatibility ( $\beta = 0.10, p = .556$ ), nor security ( $\beta = -0.15, p = .468$ ) were significantly associated with legitimacy of police.

For White participants, from the five dimensions explaining the variance of trust, only integrity was a significant predictor for legitimacy ( $\beta = 0.51, p = .017$ ). None of competence ( $\beta = 0.17, p = .282$ ), compassion ( $\beta = 0.16, p = .399$ ), compatibility ( $\beta = 0.19, p = .139$ ), nor

**Figure 2.** Parallel Mediation Analysis



<sup>1</sup> We started with the underlying dimension of security, then focused on the trustworthiness dimensions (integrity & competence), and later the relationship relevant

dimensions (compassion & compatibility). Order of inclusion did not change the patterns for either group.

**Table 3.** OLS path analysis for the indirect effects of race on legitimacy through the five trust dimensions.

Predictor	Mediator	<i>a</i>	<i>a<sub>se</sub></i>	<i>LCI</i>	<i>UCI</i>	Outcome	<i>b</i>	<i>b<sub>se</sub></i>	<i>LCI</i>	<i>UCI</i>
Race	Competence	1.28**	.20	0.88	1.68	Legitimacy	.06	.13	-0.16	2.00
	Integrity	1.39**	.20	1.00	1.78		.29	.13	-0.04	0.63
	Compassion	1.63**	.21	1.23	2.05		.47**	.15	0.17	0.77
	Compatibility	1.26**	.23	0.81	1.73		.14	.11	-0.07	0.36
	Security	1.70**	.24	1.23	2.18		-1.0	.13	-0.36	0.16

\*\*p < .001; \*p < .05; Bootstrapped 5000, Seed = 160730

**Table 4.** Multiple linear regression analysis for Black and White participants legitimacy of police.

<i>Black Participants</i>						
	<i>B</i>	<i>se<sub>β</sub></i>	<i>t</i>	<i>LCI</i>	<i>UCI</i>	
Competence	-0.11	0.20	-0.53	-0.51	0.29	
Integrity	0.13	0.26	0.49	-0.39	0.64	
Compassion	0.74*	0.23	3.16	0.27	1.20	
Compatibility	0.1	0.17	0.59	0.27	1.20	
Security	-0.15	0.20	-0.73	-0.58	0.26	
<i>White Participants</i>						
Competence	.065	0.15	1.08	-0.14	0.47	
Integrity	0.51*	0.21	2.44	0.09	0.92	
Compassion	0.16	0.19	0.85	-0.14	0.47	
Compatibility	0.19	0.13	1.49	-0.06	0.54	
Security	-0.09	0.17	-0.56	-0.42	0.24	

Note: dependent variable: legitimacy of police; β = unstandardised beta coefficient; 95% LCI and UCI; \* p < .05; \*\* p < .01

security (β = -0.09, *p* = .576) were significantly associated with legitimacy of police.

Overall, these findings suggest that Black participants lower views of police as a legitimate institution are best predicted by the lack of compassion-based trust, whilst White participants views of police as a more legitimate institution are best predicted by integrity-based trust. In a general sense, the lack of compassion-based trust evident with Black participants suggests that the Black community don't trust that the police care about Blacks wellbeing, and this lack of compassion-based trust undermines the perceived legitimacy of the police. At the same time, White participants views of police as a more legitimate institution is best predicted by integrity-based trust – the perception that the police are honest and act in accordance with moral codes. We will return to these findings in regards to different policing approaches and their impact on minority communities in the discussion. In conclusion, our findings indicate two different processes for perceived legitimacy of minority and majority participants, importantly predicted by different dimensions of trust.

**DISCUSSION**

This study examined ethnic-based differences in perceived legitimacy, focusing specifically on recently identified distinct types of trust. First, consistent with past research (e.g., Kahn et al., 2017; Oliveira & Murphy, 2015; Tyler, 2005; Van Craen & Skogan, 2015), we

established that minority group members were found to have lower trust in police than majority group members. Importantly, this pattern of findings was also replicated for legitimacy: majority group members view the police as a more legitimate institution than minority group members do (Tyler, 2010, 2011).

Interestingly, given that there is a difference between how legitimate minority versus majority group members view police, we found no direct relationship between race and legitimacy. There was, however, an indirect effect of race on legitimacy, via compassion-based trust. This indicates that the more compassionate police are perceived to be, the more legitimate they are perceived to be. In other words, our data indicates that legitimacy hinges upon a compassion-based trust perception. Further, the relationship between compassion-based trust and legitimacy was found to be especially important with respect to minority group members. Multiple linear regression analyses suggested that minority group members reduced view of police as a legitimate institution are best predicted by compassion-based trust, whilst for majority group members legitimacy was best predicted by integrity-based trust. This revelation may have important implications for the way police interacts with communities, particularly their police approaches.



### ***Policing Approaches: Instrumental versus Trust-Based***

Broadly speaking there are two competing approaches police take. The first is an instrumental approach, which is focused on a duty to obey, such that communities comply with the police based on either threat of punishment or compliance based on obligation (Tyler, 2016). Duty to obey also aligns with the perception that the police hold the moral mandate to police the communities and do in-line with a moral value (Jackson & Gau, 2016). However, the instrumental approach seems to foster perceived legitimacy through integrity-based trust, the perception that the police is acting according to moral values and righteously. Our findings that for White Americans the legitimacy of the police is predicted by higher integrity-based trust, can indicate the instrumental approach aligns more with ethnic majority perception on policing.

Concerningly, the instrumental approach to police, such as the duty to obey tends to be disproportionately focused on minority group members (Gelman et al., 2005; Ayres & Borowsky, 2008). For example, in the early 2000s, the ‘duty to obey-approach’ experienced strong public support in New York City, and police stops (designed to act as deterrents) increased 500% (Tyler, 2011). Contrary to expectation, however, crime rate during this time did not change, suggesting such an instrumental approach (especially when focused on the ethnic minority) is ineffective in increasing cooperation with police (Fagan et al., 2009). Our findings that, for Black Americans, the legitimacy in police is predicted not by a ‘duty-to-obey’ integrity-based trust, but instead a compassion-based trust, supports this conclusion.

Therefore, a strong-suit police approach or strong “law and order” (e.g., a punitive style of policing, attempting to gain compliance via threat of punishment), where police respond with a lack of compassion, does not only destroy trust, it consequentially also reduces legitimacy, particularly where the ethnic minority is concerned. This suggests that law-and-order approaches to policing are often uncalled for, as our data indicates that this is harmful – not only for communities, but also for police insofar as this results in lower legitimacy, reducing communities willingness to cooperate with the police and their mandate (Tyler, 20050).

An alternative approach to policing is a trust-based approach, in which communities are internally motivated to engage with the police (Tyler, 2016). The trust-based approach’s mechanism of cooperation between communities and police is beyond obligation and deterrence. It takes the agency of the communities into account, going beyond a more passive rule of being policed. Rather than expecting communities to blindly follow police instruction, such trust-based approaches focus on communities cooperating with police due to internal beliefs that it is the right thing to do (Jackson & Gau, 2016). Furthermore, a trust-based approach also puts a higher obligation on the police to earn the trust of communities; the police need to be trustworthy and also work to establish trustful relations with the communities (Tyler, 2005). This aligns with the finding our work, which indicates as well that minority communities

emphasise a more relational trust, such as compassion-based trust, to enhance perceived legitimacy.

However, more often than not, a general sense of “trust in police” is implicitly understood as what the IGT-Model would deem integrity-based trust. For example, when differentiating between trust and legitimacy, Jackson and Gau (2016) discuss trust in police with respect to appropriateness of police action. In other words, societies trust in police is related to how well they feel police protect their communities’ rights, and to how just police decisions are. When comparing this view of trust to Kappmeier (2016; Kappmeier, Guenoun & et al., 2021) Intergroup Trust Model, it is most akin to the idea of integrity-based trust. Integrity-based trust here is considered the level of confidence communities have that the police will be honest, and will act in accordance with a moral code.

If people generally conceptualise “trust” in line with institutional-based trust, our findings suggest that only majority group perspectives are being considered. This may have important implications when it comes to interventions designed to increase the perspective of police legitimacy. Indeed, whilst increasing societies trust that police will behave in ways considered right and just, may enhance the perceived legitimacy of police with respect to majority group members (as supported by our findings). However, our results suggest that this strategy may be ineffectual when it comes to minority group members. In order to enhance legitimacy of police from the perspective of minority group members, we would suggest that strategies would need to be implemented the consider increasing compassion-based trust, i.e., trust that police will treat me with compassion and are concerned with the needs of my community.

Taken as a whole, our findings suggest two important take-aways. First, a strong-suit law-and-order policing approach undermines legitimacy. Second, the general “trust in police” approach considers only the majority groups perspective, perhaps further undermining legitimacy of police for the minority group. Overall, considering our findings that the dimension of trust that predicts perceived legitimacy of police differs across minority and majority group members. This conclusion could have significant implications for how police as an institution might introduce strategies to increase their perceived legitimacy.

### ***Relevance of the current findings to a New Zealand perspective***

The current research assesses which dimensions of trust predict views of police as a legitimate institution, while taking ethnic-based difference into account. Our findings emphasise the importance of considering the perspectives of both minority and majority group members, as different trust dimensions predict legitimacy of police for such groups. However, our data was sourced from a US sample, considering groups of Black and White participants. Nevertheless, we have reason to believe that a similar pattern of results may hold when considering a New Zealand cultural context.

Ethnic minority groups across both the USA (e.g., Blacks) and NZ (e.g., Māori) routinely report lower trust in police compared to majority group members (e.g.,

Whites and Pākehā respectively). Further, preliminary research by Kappmeier, Guenoun & et al., (2019) directly compared which trust dimensions of the IGT-Model capture trust in the police across minority groups in the USA (Black Americans) and NZ (Māori), when faced with discrimination. The results suggest similar patterns across both groups: perceived discrimination lowered trust in the police for both groups, which is mediated by a lack of compatibility-based trust. Importantly, from the perspective of the present research, this suggests that trust in police follows a similar pattern when considering Blacks in America or Māori in New Zealand. Though not explicitly tested as of yet, this leads us to believe that it is possible for the relationship between police legitimacy and minority group status, mediated via particular dimensions of trust, to also be mirrored across these cultural contexts. Nevertheless, further research investigating this relationship is needed.

Whilst we have some evidence to suggest similar dimensions of trust in police are important from Māori versus Blacks perspectives (e.g., Kappmeier, Guenoun & et al., 2019), we cannot yet confidently conclude that this is indeed the case. Indeed, policing does not occur in a vacuum, and any considerations of contemporary policing must always consider past experiences (Kappmeier &

Mercy, 2019). Whilst Māori and Blacks may report similar histories with police in terms of discrimination and unjust treatment (e.g., Quince, 2007; Howell et al., 2004; Te Whaiti & Roguski, 1998), the details of how this was experienced differs across these groups.

Therefore, an important next step of this research is to consider the relationship between trust in police and perceived legitimacy of police from a New Zealand cultural context, e.g., across Māori and Pākehā groups. Current research of the first author is undertaking this endeavour, but at this point no empirical data has been provided yet.

Our research and results illustrate the specific facet of trust that can help to explain ethnic-based differences in trust and suggest that ethnic minority and majority members focus on different components of trust when evaluating the legitimacy of the police. Given the observed inequality, with a disproportional higher arrest number and incarceration of Māori, we need to find a way to address the shortcomings of our justice system. Understanding the basis for trust and legitimacy, and pathways for building trust in (and therefore increasing perceived legitimacy of) police is a first imperative step towards this end.

## References

- Ayres, I., & Borowsky, J. (2008). *A Study of Racially Disparate Outcomes in the Los Angeles Police Department*. Report to the ACLU of Southern California.
- Balliet, D., & Van Lange, P. A. M. (2013). Trust, Conflict, and Cooperation: A Meta-Analysis. *Psychological Bulletin*, 1090-1112. doi: 10.1037/a0030939
- Burgason, K. A. (2017). Trust is hard to come by: Disentangling the influence of structural characteristics and race on perceptions of trust in the police in a major Southern County. *Journal of Ethnicity in Criminal Justice*, 15(4), 349-371. doi: 10.1080/15377938.2017.1385558
- Connelly, B. L., Crook, T. R., Combs, J. G., Ketchen, D. J., & Aguinis, H. (2018). Competence- and Integrity-Based Trust in Interorganizational Relationships: Which Matters More? *Journal of Management*, 44(3), 919-945. doi: 10.1177/0149206315596813
- Fagan, J., Geller, A., Davies, G., & West, V. (2009). Street stops and broken windows revisited: The demography and logic of proactive policing in a safe and changing city. In S. K. Rice & M. D. White (Eds.), *Race, Ethnicity, and Policing*. New York University Press.
- Field, A. P. (2013). *Discovering Statistics using SPSS: And Sex and Drugs and Rock n'Roll* (3rd ed.). SAGE Publications.
- Gelman, A., Kiss, A., & Fagan, J. (2005). *An analysis of the NYPD's Stop-And-Frisk Policy in the context of claims of racial bias*. Columbia Public Law Research Paper, 05-95.
- Griffiths, C. E. (2018). The disjuncture between confidence and cooperation: Police contact amongst Polish migrants and established residents. *European Journal of Criminology*, 15(2), 197-216. doi: 10.1177/1477370817712962
- Howell, S. E., Perry, H. L., & Vile, M. (2004). Black cities/White cities: Evaluating the police. *Political Behaviour*, 26(1), 45-68. doi: 10.1023/B:POBE.0000022343.51330.1e
- Jackson, J., Bradford, B., Stanko, B., & Hohl, K. (2013). *Just Authority? Trust in the police in England and Wells*. Routledge.
- Jackson, J., & Gau, J. M. (2016). CarAving up concepts? Differentiating between trust and legitimacy in public attitudes towards legal authority. In E. Shockley, T. M. S. Neal, L. M. PytlikZillig, & B. H. Bornstein (Eds), *Interdisciplinary Perspectives on Trust* (p. 49-69). Springer International Publishing.
- Jackson, J., Huq, A. Z., Bradford, B., & Tyler, T. R. (2013). Monopolizing force? Police legitimacy and public attitudes toward the acceptability of violence. *Psychology, Public Policy, and Law*, 19(4), 479-497. doi: 10.1037/a0033852
- Kääriäinen, J., & Niemi, J. (2014). Distrust of the Police in a Nordic Welfare State: Victimization, Discrimination, and Trust in the Police by Russian and Somali Minorities in Helsinki. *Journal of Ethnicity in Criminal Justice*, 12(1), 4-24. doi: 10.1080/15377938.2013.819059
- Kahn, K. B., Lee, J. K., Renauer, B., Henning, K. R., & Stewart, G. (2017). The effects of perceived phenotypic racial stereotypicality and social identity threat on racial minorities' attitudes about police. *The Journal of Social Psychology*, 157(4), 416-428. doi: 10.1080/00224545.2016.1215967
- Kappmeier, M. (2016). Trusting the Enemy—Towards a Comprehensive Understanding of Trust in Intergroup Conflict. *Peace and Conflict: Journal of Peace Psychology*, 22(2), 134-144. doi:10.1037/pac0000159
- Kappmeier, M., Guenoun, B., & Campbell, R. (2019). They Are Us? The mediating effects of compatibility-based trust on the relationship between discrimination and overall trust. *New Zealand Journal of Psychology*, 48(1), 91 - 105.
- Kappmeier, M., Guenoun, B., & Fahey, K. H. (2021). Conceptualizing trust between groups: An empirical validation of the five-dimensional intergroup trust model. *Peace and Conflict*, 27(1), 90-95. doi:10.1037/pac0000537

- Kappmeier, M., & Mercy, A. (2019). The long road from cold war to warm peace: Building shared collective memory through trust. *Journal of Social and Political Psychology, 7*(1), 525-555. doi:10.5964/jsp.p.v7i1.328
- Kappmeier, M., Venanzetti, C., & Inton-Campbell, J. M. (2021). No Peace without Trust: The Trust and Conflict Map as a Tool for Reconciliation. In K. P. Clements & S. Lee (Eds.), *Multi-Level Reconciliation and Peacebuilding* (pp. 110 - 134). New York, NY, US: Routledge/Taylor & Francis Group.
- Kelman, H. C., & Hamilton, V. L. (1989). *Crimes of obedience: Toward a social psychology of authority and responsibility*. Yale University Press.
- McEvily, B., & Tortoriello, M. (2011). Measuring trust in organisational research: Review and recommendations. *Journal of Trust Research, 1*(1), 23-63. doi: 10.1080/21515581.2011.552424
- Miles-Johnson, T., & Pickering, S. (2018). Police recruits and perceptions of trust in diverse groups. *Police Practice and Research, 19*(4), 311-328.
- Moravcová, E. (2016). Willingness to cooperate with the police in four central European countries. *European Journal on Criminal Policy and Research, 22*(1), 171-187. doi:10.1007/s10610-015-9271-0
- Mummolo, J. (2018). Militarization fails to enhance police safety or reduce crime but may harm police reputation. *Proceedings of the National Academy of Sciences of the United States of America, 115*(37), 9181-9186. doi: 10.1073/pnas.1805161115
- Murphy, K., Cramer, R. J., Waymire, K. A., & Barkworth, J. (2018). Police bias, social identity, and minority groups: A social psychological understanding of cooperation with police. *Justice Quarterly, 35*(6), 1105-1130. doi:10.1080/07418825.2017.1357742
- Oliveira, A., & Murphy, K. (2015). Race, social identity, and perceptions of police bias. *Race and Justice, 5*(3), 259-277. doi:10.1177/2153368714562801
- Panditharatne, S., Chant, L., Sibley, C. G., & Osborne, D. (2018). At the Intersection of Disadvantage: Socioeconomic Status Heightens Ethnic Group Differences in Trust in the Police. *Race and Justice, 2153368718796119*. doi: 10.1177/2153368718796119
- President's Task Force on 21st Century Policing. (2015). *Final Report of the President's Task Force on 21st Century Policing*. Office of Community Oriented Policing Services.
- PytlikZillig, L. M., & Kimbrough, C. D. (2016). Consensus on conceptualizations and definitions of trust: Are we there yet? . In E. Shockley, T. M. S. Neal, L. M. PytlikZillig & B. H. Bornstein (Eds.), *Interdisciplinary perspectives on trust : towards theoretical and methodological integration* (pp. 17-47): Cham : Springer.
- Quince, K. (2007). Māori and the Criminal Justice System. In J. Tolmie (Ed.), *Criminal Justice in New Zealand*. LexisNexis NZ.
- Sargeant, E., Murphy, K., & Cherney, A. (2014). Ethnicity, trust and cooperation with police: Testing the dominance of the process-based model. *European Journal of Criminology, 11*(4), 500-524. doi: 10.1177/1477370813511386
- Schuck, A. M., Rosenbaum, D. P., & Hawkins, D. F. (2008). The influence of race/ethnicity, social class, and neighborhood context on residents' attitudes toward the police. *Police Quarterly, 11*(4), 496-519.
- Sivasubramaniam, D., & Goodman-Delahunty, J. (2008). Ethnicity and Trust: Perceptions of Police Bias. *International Journal of Police Science & Management, 10*(4), 388-401. doi: 10.1350/ijps.2008.10.4.094
- Te Whaiti, P., & Roguski, M. (1998). *Māori Perceptions of the Police*. Retrieved from <http://www.police.govt.nz/about-us/publication/police-perceptions-Māori-1998>
- Tyler, T. R. (2005). Policing in black and white: Ethnic group difference in trust and confidence in police. *Police Quarterly, 8*(3), 322-342.
- Tyler, T. R. (2006). Psychological Perspectives on Legitimacy and Legitimation. *Annual Review of Psychology, 57*, 375-400.
- Tyler, T. R. (2010). Legitimacy in corrections: Policy implications. *Criminology and Public Policy, 9*(1), 127-134.
- Tyler, T. R. (2011). Trust and legitimacy: Policing in the USA and Europe. *European Journal of Criminology, 8*(4), 254-266.
- Tyler, T. R. (2016). Trust in the Twenty-First Century. In E. Shockley, T. M. S. Neal, L. M. PytlikZillig & B. H. Bornstein (Eds.), *Interdisciplinary perspectives on trust: Towards theoretical and methodological integration* (pp. 203-216). Springer.
- Tyler, T., & Cremer, D. D. (2009). Ethics and rule adherence in groups. In D. De Cremer & D. De Cremer (Eds.), *Psychological perspectives on ethical behavior and decision making* (pp. 215-232). Information Age Publishing.
- Tyler, T. R., & Huo, Y. J. (2002). *Trust in the Law: Encouraging Public Cooperation with the Police and Courts*. Russell Sage Foundation.
- Tyler, T. R., & Jackson, J. (2014). Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. *Psychology, Public Policy, and Law, 20*(1), 78-95. doi:10.1037/a0034514
- Tyler, T. R., & Jeffrey, F. (2008). Legitimacy and Criminal Justice: Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities? *Ohio State Journal of Criminal Law, 6*, 231-803.
- Van Craen, M., & Skogan, W. G. (2015). Differences and similarities in the explanation of ethnic minority groups' trust in the police. *European Journal of Criminology, 12*(3), 300-323. doi: 10.1177/1477370814535375
- Wang, F., & Arnold, M. T. (2008). Localized income inequality, concentrated disadvantage and homicide. *Applied Geography, 28*(4), 259-270. doi: 10.1016/j.apgeog.2008.07.004
- Weine, S., Kohrt, B. A., Collins, P. Y., Cooper, J., Lewis-Fernandez, R., Okpaku, S., & Wainberg, M. L. (2020). Justice for George Floyd and a reckoning for global mental health. *Global Mental Health, 7*.

#### Corresponding Author

Mariska Kappmeier

Email: [mariska.kappmeier@otago.ac.nz](mailto:mariska.kappmeier@otago.ac.nz)

National Centre for Peace and Conflict Studies

University of Otago, 262 Leith Walk

9016 Dunedin

# Space for you and your baby: Participant perceptions of community-based postnatal parenting support and adjustment to parenthood

Linde-Marie Amersfoort<sup>1</sup> and Myron Dean Friesen<sup>2</sup>

<sup>1</sup>School of Health Sciences, University of Canterbury

<sup>2</sup>School of Educational Studies and Leadership, University of Canterbury

Space for You and Your Baby is a preventative support programme for new parents based on the Australian supported playgroup model. In Aotearoa New Zealand, Space is provided to approximately 2000 participants each year but has never been formally evaluated. This study employed a cross-sectional retrospective research design and examined why new parents attend Space and how Space contributed to their adjustment to parenthood. Over 500 current and former participants completed a mixed-methods survey. The results showed that participants were primarily motivated to attend Space for social support and highly endorsed the programme across all of the targeted outcomes. Facilitator competency moderated these generally positive findings. The results have implications for facilitator training, community partnerships, and point to opportunities for further evaluation research.

**Keywords:** *First-time parents, parenting support, postnatal care, programme evaluation*

## INTRODUCTION

There is strong evidence in both the quantitative (e.g., Kunseler et al., 2014; Parfitt & Ayers, 2014) and qualitative (e.g., De Haan, 2016; Wilkins, 2006) literature that, overall, parents find the transition to parenthood to be both a time of delight and joy, but also a time fraught with new challenges and stress as they adjust to the needs of their new baby and their parenting role. Thus, the transition to parenthood is deemed a period where first-time parents are at higher risk for experiencing distress and, consequently, for developing mental health difficulties (Parfitt & Ayers, 2014; Sanders et al., 2014). Even though most new parents are able to cope with the many changes that accompany their new roles, Sanders et al. (2014) found many new parents felt underprepared, unsure, alone, and inadequate as they entered into parenthood - much of which could have been prevented with greater preparation, more realistic expectations, and high-quality support both pre-and postnatally.

There is a growing body of evidence for the efficacy of parenting programmes that offer support to new parents, including the Incredible Years Parents and Babies™ Program (Webster-Stratton, 2008) and Baby Triple P (Spry et al., 2010). These have been shown to be efficacious in their aims to provide new parents, especially those considered at-risk, with specific knowledge and skills that result in improving positive parenting practices and reducing child behaviour problems over time (e.g. Evans et al., 2015; Jones et al., 2016). However, there is a paucity in the research as to the effectiveness of more universal preventative programmes focused on assisting first-time parents through parent education and support (Hickey, 2019; McLean et al., 2017). The research is also scant as to the efficacy of such preventative initiatives when delivered outside of formal

or clinical settings but within the voluntary and community sector (Gardner & Woolgar, 2018). This is despite studies suggesting that vulnerable and isolated parents are less likely to engage with formal, top-down parenting training programmes, and are more likely to engage with programmes conducted in familiar and/or informal locations, delivered by facilitators known to parents, and where social networks can offer comfort and security (Gardner & Woolgar, 2018; McLean et al., 2017).

Despite the lack of research into the efficacy of parent education and support programmes delivered in volunteer/community settings, evidence for the efficacy of parent education and support delivered through the 'Playgroup' model is growing. Research has found that playgroup attendance provides families with increased social support and connection, increased caregiver knowledge and awareness of the benefits of early childhood education, and promotes young children's social interaction (McLean et al., 2020; Mize & Petit, 2010). Whilst playgroups do not focus specifically on new parents, they can be described as groups regularly attended by caregivers and their preschool children (aged 0-5) in order to provide children with social experiences through play and parents with "child-rearing guidance and social support" (Mize & Pettit, 2010, p. 1271). Consequently, playgroups operate on the principles of play-based learning, social interaction, peer support, and positive modeling of parenting practices (Wright et al., 2019). Playgroups are especially popular in England and Australia where they are run and coordinated by local or regional bodies subject to, and supported by, a national organisation such as Playgroup Australia (McLean et al., 2020; Mize & Petit, 2010). In New Zealand, playgroups are endorsed by the Ministry of Education, and have been recognised as providing learning environments that are

varied and responsive to individual children's interests and learning needs, and informal support networks for caregivers (Ministry of Education New Zealand, n.d.).

Whilst there are many different types of playgroups in New Zealand, for example those focused on a specific culture (e.g., Pasifika playgroups), language (e.g., Ngā Puna Kōhungahunga; Māori language playgroups), or philosophical approach (e.g., Montessori playgroups), these groups are all run by the parents and caregivers themselves (Ministry of Education, New Zealand, n.d.). However, in Australia, the playgroup model offers a two-tiered approach which classifies playgroups either as community playgroups or supported playgroups (McLean et al., 2020). Community playgroups are akin to those endorsed by the New Zealand Ministry of Education in that they are caregiver-led, occur throughout a range of communities, and are attended by caregivers and children from a variety of socioeconomic and cultural backgrounds (McLean et al., 2017; McLean et al., 2020). Although supported playgroups still operate on the basic playgroup principles, in general they are not parent-led but run by trained facilitators, most often early childhood educators, and often have a greater focus on specific cohorts of families, including young parent families, migrant families, and families with children who have developmental disabilities (Commerford & Robinson, 2016; Wright et al., 2019). Jackson (2011) found the benefits of supported playgroups for parents to include friendship and social network support, peer support, emotional support, parenting role support, information and resource support, and multidisciplinary support (i.e., the opportunity to have professionals attend the playgroup and offer insights and access to services that would not have ordinarily been available to parents in informal, non-clinical settings).

The benefits of attending supported playgroups appear to align with the research concerning the needs of new parents, and mothers in particular. This is particularly true when this form of parent education is sensitive to the social context of early parenting, as well as the dramatic lifestyle changes that confront new parents, simultaneously enabling them to learn, integrate, and intuitively apply positive parenting practices (Copeland & Harbaugh, 2019; Wilkins, 2006). Thus, parent education and support that alleviate feelings of self-doubt and isolation, help parents to manage unrealistic expectations of parenthood, and offer opportunities for skill acquisition within a supportive and reassuring environment, are likely to be the most impactful (Hanna et al., 2002; Sanders et al., 2014; Wilkins 2006).

### **Space for You and Your Baby (Space)**

Space for you and your baby (Space) is a parent education and support programme focused on supporting parents during the transition to parenthood. It operates in a manner similar to that of the Australian supported playgroups. Space was developed in 2003 as a Playcentre New Zealand programme, but has now grown to include providers from a wide range of early childhood centres, as well as in settings such as community centres and churches. In order to run the Space programme, all partner organisations, including Playcentre, pay an annual licensing fee to Parenting Place. Established in 1993,

Parenting Place is a for-purpose charitable trust that develops parenting programmes, including its flagship Toolbox and Building Awesome Whānau courses. The programmes are delivered to parents nationwide within Aotearoa New Zealand through a range of community partner organisations such as churches, early childhood education centres, and social services agencies.

The Space programme sessions are attended by both the parents or caregivers and their infants. Whilst the programme can be attended by multiple caregivers and whānau, overwhelmingly the participants are the biological mothers of the infants and the programme is seldom attended by both parents, multiple caregivers, or extended whānau. The sessions are organised and delivered by trained Space facilitators. Although the majority of facilitators are early childhood educators, they also come from a range of backgrounds and include social workers and community volunteers. Parents who attend the Space programme through Playcentre register for the course directly through the Playcentre booking system for a one-off fee of \$105 which covers both their registration fee and their attendance dues. Parents who attend the programme at any other provider pay a registration fee of \$32 to Parenting Place and an additional cost to the Space partner delivering the programme. This additional cost varies amongst partner organisations and can range from \$1-\$2 per session to a termly fee of \$50. The standard Space programme involves 30-40 weekly sessions that vary between 1.5 and 2.5 hours in length. The first half of the curriculum, delivered across the first 20 weeks of the programme, has a strong parent education focus where topics such as 'Becoming a Parent', 'Infant Sleep', and 'Infant Brain Development' are presented by a facilitator or guest speaker and discussed by parents. The Space sessions include opportunities for social interaction and discussions between parents, as well as specific opportunities for infant-parent interaction through the inclusion of music, stories, and heuristic play baskets. The second half of the programme (i.e., approximately the last 20 weeks of the programme) has a stronger focus on infant-parent play. These sessions provide an opportunity for dyads to explore a new play experience through, for example, the use of elements or materials such as sand or water, or through a play modality such as 'messy play'. The play activities are organised by the facilitator and parents are encouraged to both observe and participate in their children's play experiences. During the play-based sessions, facilitators continue to support interactions and discussions relevant to the session amongst parents and also provide parents with ideas and activities that can be implemented or replicated at home. The Space curriculum incorporates some bicultural elements including opening and closing karakia (prayers) as well as karakia mō te kai (blessing of the food), whakataukī (Māori proverbs), and waiata (songs) in te reo Māori.

### **Present Study**

A literature search revealed a paucity in research examining the efficacy of parent education and support delivered through the Australian supported playgroup model that targets the transition to parenthood specifically, and the Space programme has not been previously evaluated. As part of a research and evaluation

collaboration between the University of Canterbury and Parenting Place, we worked with key programme staff to develop a theory of change (ToC) model to guide programme evaluation and redevelopment strategies (Amersfoort et al., 2021). A ToC is similar to a programme logic model, but goes further in the conceptualisation of how and why programme strategies lead to specific short- and long-term outcomes, while explicitly including the assumptions on which a programme is based, and the additional variables that may moderate the process (De Silva, Breuer, et al., 2014; Centre on the Developing Child, n.d.). Consequently, the purpose of this study was to test key assumptions and the targeted outcomes from the Space ToC through a large retrospective investigation of the programme-related experiences of current and former Space participants.

The key research questions for this study included:

1. What are the main reasons parents choose to participate in Space?
2. How do Space participants feel about how the programme promoted their development as a new parent in terms of (a) developing quality parent-infant interactions, (b) growing in parenting confidence, (c) experiencing social support and a sense of community, and (d) relating the information and strategies provided by Space to their family situation?
3. Were there significant differences in parents' self-reported Space outcomes across sociodemographic characteristics?
4. What aspects of Space did parents find most/least helpful in their journey as a new parent, and what are the main reasons parents were satisfied/dissatisfied with their participation in Space?

**METHOD**

**Participants**

The majority of survey respondents were biological mothers to the infants in their care (see Table 1). Only one participant indicated a different relationship, and only three men responded to the survey. Table 1 also shows that participants tended to be from more recent Space cohorts (2016-2018), and were slightly over thirty years of age when they first became a parent. When parents first started Space their infants ranged in age from less than one month old to nine months old (70% reported beginning Space when their infants were between two and four months old). A majority of survey respondents identified with a European/Pākehā ethnicity with far fewer numbers from Asian, Māori, and people groups from the Pacific and other geographic regions (e.g., Latin America and Africa; 3%). The majority of survey respondents were well-educated, with just over three out of four having earned a Bachelor's degree or higher level of qualification. Additionally, over 70% had returned to work, with over three quarters of these (77%) returning to occupations that, based on the 2013 Australian and New Zealand Standard Classification of Occupations, were managerial or professional (Statistics NZ, 2020).

**Materials**

*Participant engagement measures:* A mix of questions with categorical response options and open response options queried how respondents came to know about the Space programme, where they attended the programme (i.e., geographic region, community partner), and the length of the course they attended. In order to understand participants' reasons for attending a parent support programme during the transition to parenthood, we asked participants to "Briefly describe why you decided to attend a formal parenting support and education programme in the first instance" and "Briefly describe why you chose the Space programme".

*Outcomes measures from Space participation:* As described above, the outcome measures were drawn from the Space ToC. These were measured by asking participants to reflect how Space affected their feelings about being a parent, their relationship with their baby during the first year, the effectiveness of the information and strategies provided by Space that they applied in the care of their baby, and the social support network developed through Space. This section of the survey included a combination of questions drawn from established measures and adapted for the context of the current evaluation along with custom written items specific to the Space programme. All questions were scored on 5-point Likert scales (1 = Disagree; 5 = Agree). All of these items were included in a principal components analysis (PCA) with Promax rotation and Kaiser normalization due to the potential of correlated components. The first analysis identified four factors with an Eigen value greater than one. However, inspection of the Scree plot pointed to a three-factor solution, and the pattern matrix revealed items that had high loadings across factors (> .30) and a few items that did not

**Table 1.** Demographic characteristics of study participants

Demographic characteristics	M (SD) /Freq (%)	Min – Max
Gender		
Female	549 (99.5%)	
Male	3 (0.5%)	
Year of Space Participation		
2018	190 (34.1%)	
2017	148 (26.5%)	
2016	104 (18.6%)	
2015	52 (9.3%)	
2014	64 (11.5%)	
Age first became a parent	31.84 (4.44)	19 - 47
Age of infant when starting Space (months)	3.02 (1.41)	< 1 – 9
Ethnicity		
European/Pākehā	469 (85.0%)	
Asian	36 (6.5%)	
Māori	31 (5.6%)	
Pacific peoples and all other groups	16 (2.9%)	
Educational qualifications		
Tertiary degree (Bachelor's or higher)	457 (75.7%)	
Tertiary certificate or diploma	56 (10.2%)	
NCEA Level 3 or other secondary qual.	22 (4.0%)	
NCEA Level 1 or 2 (or equivalent)	9 (1.7%)	
No educational qualifications	2 (0.4%)	
Returned to employment after Space	402 (73.6%)	

NOTE: Study N = 564

Missing data across individual demographic questions ranged from 2 to 18

substantially load on any factor. After removing these items and completing a second analysis, the pattern matrix showed that all items loaded cleanly across three factors which are described below (each item loaded  $>.45$  on its respective factor and together the three factors accounted for 61% of the total variance across items). Composite scales were created by averaging the individual items from each factor.

Parents' reflections of how Space contributed to the quality of parent-infant interactions and parenting confidence included 11 items ( $\alpha = .93$ ). Sample items included: The Space programme helped me to develop a stronger bond with my baby; The Space programme helped me be more affectionate with my baby; The Space programme helped me be more effective in meeting my baby's needs; The Space programme supported me to be more confident in my parenting.

Ten items assessed the relevance and applicability of information and caregiving strategies ( $\alpha = .91$ ) facilitated by Space. Of the 10 items, six focused on the extent to which participants felt they were presented with, and able to learn, new information and caregiving strategies during the Space sessions (e.g., The Space programme facilitators supported me to learn the parenting information and strategies they provided). The remaining four items were more strongly focused on the extent to which the information and strategies they had learned during the Space sessions were relevant and applicable to their unique family situation, including their cultural heritage (e.g., The information and strategies provided by the Space programme fit well with the values of my cultural heritage).

Finally, four items assessed how parents felt about the social support network ( $\alpha = .79$ ) that Space participation provided. Items queried how Space facilitated the development of supportive friendships within Space, confidence to build supportive relationships outside of Space, understanding of community resources and supports, and sense of preparedness to access support services if required.

*Participant satisfaction measures:* Open-response items queried if participants completed their Space course and also the reasons for why some participants may not have completed the programme. Additionally, one open-response question was used to examine which aspects of the Space programme participants found most helpful as new parents (i.e., "Overall, what did you experience at the Space programme that was the most helpful for you as a new parent?").

*Formative evaluation measures:* In order to investigate which aspects of the Space programme participants felt were most and least effective in supporting them during their first year of transition to parenthood, we asked survey respondents to rank eight features of the programme according to which they valued most (1) to least (8). These programme features included: (a) the interactions with facilitators, (b) infants interacting with other infants, (c) the overall length of the programme; (d) interaction with other parents/caregivers of babies; (e) the various activities during the Space sessions; (f) the topics discussed during the Space sessions; (g) guest speakers from the community; and (h) the atmosphere of the Space sessions. Two open-response

follow-up questions asked, "Reflecting back on the Space programme, do you remember any topics that were particularly helpful, or that addressed a specific need you had at the time? Please describe the topic and how it helped you." and "Reflecting back on the Space programme, do you feel there were any important topics that were missing or not covered well enough? If so, please describe/explain."

### **Procedure**

Current and former Space participants were recruited in June of 2019 via email to respond to an online survey hosted by the University of Canterbury's Qualtrics survey platform. The survey was organised according to the key assumptions and outcomes from the Space ToC and involved a combination of quantitative (i.e., Likert scales) and open-response qualitative questions. Initial emails were sent to over 10,000 addresses, representing those who participated in a Space programme from 2014 to 2018. A total of 871 people accessed the questionnaire. Adequate responses across key variables ranged from  $n = 685$  (e.g., geographic region of Space programme) to  $n = 181$  (qualitative open-response questions).

### **Data analysis**

*Missing Data Analysis:* In order to be as inclusive as possible across participants, missing data for the quantitative variables were analysed separately for the outcome measures and the rank-order formative evaluation items. For the outcome measures, after removing all participants who failed to complete the majority of items across the four scales, the sample was reduced to  $n = 567$ . A test of missing data (Little's Completely at Random (MCAR) test) was statistically significant ( $\chi^2 = 377.19$ ,  $DF = 326$ ,  $p = .03$ ). A visual inspection of the data identified three cases each with three missing values (although on different items). Omitting these participants from a second MCAR test revealed a non-significant result ( $\chi^2 = 260.33$ ,  $DF = 264$ ,  $p = .55$ ). Since the missing values from these participants were each in separate scales, they were retained in the analyses, and other individual missing values were not replaced as composite scores for each of the measures were calculated as the average (rather than the sum) across items. A similar procedure for the rank-order items resulted in a refined sample size of  $n = 549$ , with six of these participants all having a single missing value. A visual inspection of this data revealed that for these participants one of the eight rank items had been omitted and these were individually replaced (e.g., rankings included 1, 2, 3, 5, 6, 7, 8, but no 4, so the missing descriptor was replaced with a 4).

*Quantitative Data Analysis:* The Statistical Package for the Social Sciences (SPSS) 25 was used to analyse the quantitative data. Descriptive statistics (means, standard deviations, and percentages) were calculated on all variables. Bivariate correlations (Spearman's rho or Pearson) were employed to test the associations between Space outcomes and demographic characteristics measured on ordinal or continuous metric (e.g., education and age), respectively. Analysis of variance (ANOVA) examined the possibility of mean group differences on the Space outcomes across categorical demographic characteristics (e.g., ethnicity), and variables developed



from themes identified in the qualitative data analysis. Rank-order data was analysed with the non-parametric Friedman Test and Wilcoxon signed-rank post-hoc analyses with a Bonferroni correction ( $p < .006$ ).

*Qualitative Data Analysis:* The second author and a research assistant coded the data to individual qualitative questions, with each person analysing one or more specific questions. Both coders had access to the full range of data, trained together, and frequently reviewed each other's coding strategies to ensure similar strategies and coding schemes were being applied. Due to the collaborative nature of the qualitative coding strategy, formal estimates of inter-rater reliability were not assessed. A combination of content analysis and thematic analysis was applied to the data. First, content analysis was applied to participant responses that included specific references to course content and programme elements. Then, following the procedures recommended by Braun and Clarke (2006), thematic analysis was employed to examine those responses that were more generic and to organise all coded data into the broader themes.

**RESULTS**

**Space participation**

Participants were referred to Space from a variety of sources. The majority of parents (56%) discovered Space through a word-of-mouth referral, followed by referrals from antenatal groups (18%), postnatal well-health providers (11%), and midwives (10%). Very few (<5%) found Space through advertising or an internet search. Overall, survey respondents primarily participated in Space programmes that were delivered in the major centres of the North Island of New Zealand (88.4%), with 32.6% of all survey respondents attending Space programmes that were delivered in Auckland. Of the small cohort of survey respondents who attended programmes in the South Island (11.5%), the greatest proportion attended their Space programmes in Dunedin (5.4% of all survey respondents). Reflecting the historical roots of Space, 69.7% of participants attended Space through their local Playcentre, followed by church denominations (12.2%), and then other Early Childhood Education providers (9.1%). The remaining participants (9%) attended Space through a variety of other social service/community providers.

Well over three quarters of survey respondents participated in either a 40-week Space programme (53.8%) or a 30-week Space programme (28.9%). Attendance in programmes with a duration of 25 or 20 weeks were much less common (3.8% and 8.1%,

respectively). Over three quarters of all participants indicated that they completed the full Space programme that was offered. Those who did not complete a full programme ( $n = 143$ ) identified five broad issues: (a) returned to work (36%); (b) personal or baby-related circumstances (21%); (c) still attending a course (17%); (d) course content or facilitator issues (14%); and (e) enrolled in the course at a late stage (8%).

**Motivation to attend Space**

The majority of survey participants (71%) indicated that they decided to attend a formal parenting programme because they wanted to meet other parents, make friends, and receive support as a first-time parent with others having the same experience. Three additional reasons that were listed by almost 10% or more of respondents included a motivation to learn about parenting and child development during the infant years (26%); a need to get out of the house (13%); and the desire to engage in a structured activity with their infant (9.5%).

Most parents chose Space specifically because it was recommended by, or they actually attended with, someone they trusted (52%); because it was conveniently located or the only programme of its kind in their area (17%); it was educational, interesting or addressed their parenting needs (16%); they heard positive reviews (13%); it was an opportunity to meet other parents with similar aged children (10%); and it was affordable (7%). Most respondents (72%) also indicated that they participated in other parenting group activities with their infants in the first year. The vast majority of these (79%) included informal parent-led activities (e.g., antenatal coffee groups), followed by miscellaneous baby and toddler activities such as baby massage or baby sensory classes (39%), swimming lessons (29%), a formalised music programme for babies called Mainly Music (27%), other organised parenting support programmes (25%), and library activities (e.g., story time; 13%).

**Outcomes from Space participation**

As per Table 2, the vast majority of participants felt that Space successfully contributed to all of the outcomes under investigation, with over 85% of respondents scoring above the "Neutral" midpoint of the five-point scale. Participants were most favourable about how Space provided relevant and applicable information and caregiving strategies, followed by the support for positive parent-infant interaction and parenting confidence, and finally the facilitation of a social support network.

*General aspects of the Space programme valued by participants:* Table 3 shows the list of the eight Space

**Table 2.** Parent related outcomes from the Space theory of change

Variable	M (SD)	% of M Scores <2	% of M Scores 2-3	% of M Scores >3-4	% of M Scores >4
Parent-infant interaction and parenting confidence	3.85 (0.79)	3.4	9.6	42.9	44.1
Relevance and application of information and caregiving strategies	4.09 (0.67)	0.9	7.6	33.0	59.6
Social support network	3.75 (0.89)	4.1	18.1	40.1	37.8

Note. N = 564

**Table 3.** Participant rank order of Space programme components from most (1) to least (8) valuable

Space programme component	Mean (SD)
Interacting with other parents/caregivers of babies	2.22 (1.73) <sup>a</sup>
My baby being able to interact with other babies	3.50 (2.12) <sup>b</sup>
The activities (e.g., music, reading books, craft-making) during the sessions	3.65 (1.89) <sup>b</sup>
The interactions with the facilitators	4.27 (1.99) <sup>c</sup>
The topics discussed	4.31 (1.80) <sup>c</sup>
The atmosphere of the sessions	4.71 (1.91) <sup>d</sup>
Guest speakers from the community	6.23 (1.94) <sup>e</sup>
The overall length of the programme	6.45 (1.89) <sup>e</sup>

Note. *N* = 549; Different superscripts in the Mean (SD) column indicate statistically significant differences (*p* < .001).

components that parents ranked from most to least important. The results clearly show that participants valued the interaction with other parents/caregivers the most, which is congruent with the most common motivating factor for attending Space. Baby interaction and the Space activities were the next most valuable components, followed by the interactions with facilitators and the various topics discussed as part of the Space curriculum. The atmosphere of the Space sessions came in sixth place, followed some distance by guest speakers and the overall length of the programme. Thus, it would seem that the survey respondents valued the social interaction and activities (which, in effect, also aid social interaction) significantly more than the educational component (i.e., the curriculum topics) of the programme.

Qualitative analyses of over 500 responses to the question about what was “most helpful” about Space showed that 94% identified some aspect of *community and connection* as the most helpful experience. The value of *community and connection* was described in terms of finding a community of similar parents with opportunities for interaction, sharing, support, encouragement, and gaining information. Within this *community and connection* macrotheme, the idea of “sharing” was mentioned by one third of survey participants. Participants valued the sharing of information and strategies, but also sharing their difficult emotions, accomplishments, and mistakes. The various ways that the experience of community was valuable for participants are illustrated in the quotes below:

*Meeting other mothers and babies that were going through the same thing and being able to support each other in a safe friendly place.*

*Connecting with other new parents and discussing how our babies were developing each week, know[ing] that it is often normal development and you are not in it alone.*

*Meeting with other new parents and sharing our highs and lows in a structured and formal way, that was monitored and facilitated.*

*The chance to learn in a relaxed, supportive, caring environment and the chance to connect with other mums and their babies and see how they did things.*

Apart from the macrotheme of *community and connection*, there were two other related categories of experiences that parents felt were also most helpful to them during Space. One in five survey respondents (20%) identified the *information* provided by facilitators and other parents as most helpful. This included general information around infant development, and more specific information on important topics such as nutrition and feeding, sleep, play, attachment, self-care, baby massage, and first-aid. One participant wrote, *Knowledge about my babies’ development and strategies to support this i.e. settling, feeding teething, physical movement, and emotions. Also enjoyed learning different songs/activities and listening to other parents and realising we all have the same struggles but at different times.*

Related to the theme of *information*, just over 10% of parents identified the activities that they and their infants did together as the most helpful aspect of Space. This included the singing and music, crafts, and creative and messy play:

*Music! We now sing and dance all the time as she loves it*

*The play term! I loved learning about the importance of play for child development and getting practical ideas for how we could have fun with my baby at home.*

*Space curriculum topics participants found helpful:* Survey participants were also asked to reflect on the extent to which particular topics, delivered as part of the Space curriculum, were helpful. Far fewer parents responded to this question (*n* = 365) compared to the aforementioned ranking question (*n* = 550). Even so, almost all of the responses could be included in seven categories. First, a majority of parents (68%) identified one or more child development topics that were particularly helpful, including neurological development, physical and motor development, emotional development, sensory development, and topics related to attachment, language, and child temperament. Second, just over half of parents (57%) also identified a topic around child safety that was helpful. This included specific topics such as water safety, general household safety (i.e., baby-proofing), CPR and first-aid, and the safe use of car seats. Following these broad themes, the two specific topics of nutrition (26%) and sleep (21%) were identified by many parents as important. Finally, parents also identified specific activities that they enjoyed, such as music, baby massage, reading, and crafts (14%), followed by the broad topic of play (11%), and then topics that encouraged parents to reflect on the process of becoming a parent and how their own lives were changing as a result of this major shift in their roles (11%).

*Space curriculum topics participants found lacking:* Survey participants also self-reported on any important topics that were missing from the Space curriculum or not covered well enough. This question had the fewest

responses of all ( $n = 181$ ). Rather interestingly, certain topics that were identified as particularly helpful by participants were also identified as needing more time and attention. These topics included those on sleep (16%) and nutrition (i.e., breastfeeding, weaning, and introducing solids; 11%). Although these topics may have been addressed by facilitators, some participants felt that the facilitators did not cover the topics adequately, were based on the opinions of the facilitators rather than being evidence-informed, or that only one perspective was presented. However, the theme that was identified by the most participants as needing further attention was maternal mental health (i.e., postnatal depression, anxiety, stress, and recovery from birth trauma; 19%). The following two examples illustrate the range of how this was described by parents:

*We did not cover or get a chance to discuss mental health of the mums. I myself was struggling with postnatal depression, but never felt safe enough to discuss it with the other mums or facilitators. It just never came up and I left it.*

*More information about mum's health - physical well-being post birth, nutrition while breastfeeding, support while sleep deprived and support with mental wellbeing (and how to access this if needed).*

*Other reasons for participant satisfaction/dissatisfaction with the Space programme:* In their responses to the question regarding missing or inadequately presented topics, almost one in ten respondents (9%) mentioned facilitator competency. Facilitator competency was also mentioned as a reason for participants stopping their attendance of the course early ( $n = 15$ ; 10.5% of those who did not complete), citing issues of disorganisation, a lack of familiarity with the curriculum, ignoring individual differences in parents' ability to cope and infants' developmental progress, pushing personal opinions in spite of other evidence, and facilitators being perceived as judgemental. Although the quantitative scores suggested that the vast majority of participants felt well-supported by their facilitators in learning and applying the information and strategies provided in the Space curriculum (83% indicated 'Agree somewhat' or 'Agree' to this question), the impact of poor facilitator competency on participants' experience is illustrated in the following quotes:

*The whole thing was just terrible. We started with about 15 families, by the end of the first term there were 3 left. The facilitators were so poorly informed and judgemental. I still get angry when I recall these sessions.*

*I don't feel like anything we covered was relevant. All our facilitator did was bully us as we didn't follow her way of parenting and three friends left the group in tears. I didn't have to go back as luckily I started back to work the week after. Otherwise I was likely to give her a piece of my mind and place a complaint.*

As mentioned above, and in contrast to these participants, many more participants expressed satisfaction and positive sentiments toward their facilitators, their competency, and their willingness to go

the extra mile in supporting their participants ( $n = 60$ ; 11% of all responses for the 'Most helpful' aspect of the course). Two examples of these positive responses are provided below:

*Our facilitators created a safe space for us to be completely honest without fear of judgement. It meant that you could vent about your struggles and hear other people's honest experiences. It helped to know that other people were often having the same struggles. They responded with empathy and gave advice when called for, or just reminded us to trust our instincts and not listen to mother/mother-in-law/sister/blogger etc.*

*My child was diagnosed with...whilst I was attending Space...The facilitators were amazing - loving, caring, supportive, arranged a care package (food!), encouraging, totally non-judgmental and keen to learn about [my child's illness]. The[ir] support and "cheering on" was the most helpful thing for me. And being able to come to Space where my baby was just one of many, ...just a baby like everyone else.*

*Individual differences:* As is shown in Table 2, between 0.9% and 4.1% of survey respondents had scores in the lower ranges of the scale for the four Space outcomes, and another 7.6% to 18.1% had scores that were lower than the midpoint of the scale. This suggests that there was a small group of parents who did not find Space to be very beneficial for them. In order to identify factors that might have contributed to the low scores of these parents, we created a dummy variable (1 = Yes, 0 = No) for all participants who either responded with frustration or disappointment over the competence of their facilitator in an open-response question ( $n = 16$ ), or ranked the interaction with facilitators in the lowest quartile of all programme components. This variable was included in a multivariate analysis of variance (MANOVA) that compared the mean scores for the Space outcomes for those who were dissatisfied with how the course was facilitated ( $n = 88$ ; 15.6% of all respondents) compared to the rest of the sample. The results showed highly statistically significant group differences with small to medium effect sizes. Those dissatisfied with how the course was facilitated provided significantly lower scores across the three Space outcomes ( $M$  differences ranged from 0.48 to 0.66,  $F$ s (1, 561) ranged from 22.33 to 59.01, all  $p$ s < .001; partial eta squared ranged from .04 to .10). Thus, it seems one explanation for why some parents reported substantially lower Space outcomes involved issues with perceived facilitator competency.

Further analyses tested if there were significant differences in the outcomes reported in Table 2 based on participants' demographic characteristics (age, educational qualification, ethnicity, and current employment status). The results showed only one statistically significant association. Space participants with higher educational qualifications reported slightly lower experiences of social support ( $r = -.10$ ;  $p = .02$ ).

## DISCUSSION

This research provides the first preliminary evidence that a supported playgroup for new parents and their infants has the potential to effectively address new

parents' needs for social support and provide a sense of community in an environment where parents can address their concerns about the personal and caregiving challenges they are facing, and receive evidence-informed education about effective caregiving practices. Although the results from this study largely support the assumptions and targeted outcomes from the Space ToC (Amersfoort et al., 2021) and compliments research on the effectiveness of supported playgroups (Commerford & Robinson, 2016; Wright et al., 2019) and the needs of new parents (De Haan, 2016; Lee et al., 2019; Sanders et al., 2014), the following discussion will address several qualifications to these findings and the opportunities provided for further research.

### **Motivation for postnatal parenting support**

The qualitative results showed that parents' major motivations for attending Space was to improve social connections with other parents who were going through the same experience, to gain a sense of community, and to receive social support. Correspondingly, participants rated the social interaction with other parents and facilitators as the most valuable component of Space. These findings echo research on the needs of new parents from broad surveys (Hanna et al., 2002) as well as in depth qualitative studies (Paris & Dubus, 2005), and supports the ToC assumption that the transition to parenthood involves a period where new parents feel isolated, and therefore, are more likely to seek out postnatal support programmes such as Space. Because the current sample was largely Pākehā, it is an open question whether this issue of social isolation is also a motivating factor for new mothers who are Māori or Pasifika and may have more established social support networks. A recent study found that expectant Māori mothers have higher rates of anxiety and depressive symptoms and life stress than non-Māori (Signal et al., 2017), but the role of social support was not investigated and is an opportunity for further research.

### **Space ToC outcomes**

In both the quantitative and qualitative results, study participants felt that Space contributed to all three of the ToC outcomes that were queried (i.e., developing quality parent-infant interactions combined with growing parenting confidence, being able to relate and apply Space content to their family situation, and experiencing social support and a sense of community), and we did not find any evidence of significant associations between the quantitative scales and participant sociodemographic characteristics. These results are difficult to compare to previous studies as there are very few supported playgroups that target infants and their caregivers (for a review see Williams et al., 2018). In a supported playgroup with mothers and children ranging from four to forty months old, Bohr et al., (2010) found reductions in parental stress and improvements in parenting confidence, but no significant changes in maternal sensitivity. As acknowledged by some parents in this study, their shifts in parenting confidence may have come naturally as they became adjusted to being a parent. How that may (or may not) be enhanced by participating in Space could only be addressed in a future study with a matched control group.

In terms of perceived social support and social networking, the findings suggest that the majority of

participants felt their Space facilitators provided excellent support in developing the social connections that they were looking for, and that the programme facilitated the development of a peer support network that provided parents with a sense of community during the time that they attended Space. The literature suggests that four elements of social support are relevant for new mothers to gain confidence in caring for their baby; informational, instrumental, emotional, and appraisal (Glavin et al., 2017; Leahy Warren, 2005). The qualitative findings showed that parents could recall receiving all four elements of social support both from their fellow participants and facilitators, but seem to have valued affirmational, emotional, and informational support more than the instrumental support.

### **Individual differences and participant reflections on Space strengths and limitations**

An important component of the Space ToC is the identification of several variables that may serve to moderate targeted outcomes, including the infrastructure and administration provided by Parenting Place in collaboration with local partner organisations, characteristics of the community partner (e.g., extent of investment into and quality of relationships developed in their local community), characteristics of participants and their infants, and most importantly, the competency and skill of the facilitators (Amersfoort et al., 2021). Our findings showed that the majority of Space participants highly valued the skills and/or support of their facilitators. However, those who felt that facilitator competency was insufficient reported significantly lower agreement that Space contributed to the four target outcomes.

This provides a challenge to community partners and Parenting Place trainers in how to assess and support the development of core facilitator competencies. Not only do facilitators have to have warmth and emotional insight and understanding of the experiences of first-time parents, but they also need to have the necessary skills to facilitate group cohesion and peer relationships, as well as the necessary knowledge regarding child development and infant caregiving. Meeting such requirements could be seen as difficult, particularly for facilitators who are paraprofessionals and/or volunteers in their community. This may point to the value of co-facilitation, where facilitators can utilize complementary strengths, and a need for facilitators to better utilize publicly available evidence-informed resources when some topics go beyond their expertise – both of which are highly-valued components of Australian supported playgroups (Jackson, 2011).

The topic most frequently mentioned as missing or needing more attention was maternal mental health. This is not unexpected since the transition to parenthood marks a time of increased vulnerability for parents (Berlin et al., 2016; Hanna et al., 2002), with increasing evidence of the influence of maternal mental health challenges on child outcomes (Howard & Khalifeh, 2020). Research also suggests that most new mothers are reluctant to seek formal help for mental health challenges (Fonseca et al., 2015; Signal et al., 2017), yet the majority of participants in this study felt that Space provided a safe and supportive environment. Thus, it would seem that the Space

programme could be a suitable context for participants to explore such an emotionally salient topic. However, as facilitator competency appears to be a crucial element in the extent to which parents may find the Space programme beneficial, alternative methods of delivery of sensitive content such as information about maternal mental health should be explored. This could happen through guest speakers who are experts in the field, or through different mediums such as carefully curated video content. Additionally, facilitators should be encouraged to connect with, and work alongside, maternal mental health organisations in their communities or, when not available, national organisations that specialise in these areas. Parenting Place should also consider providing facilitators and partner organisations with the necessary professional development to ensure that facilitators have the capacity and insight to deal with this topic sensitively and compassionately.

### **Strengths, limitations, and conclusion.**

There are a few strengths in this study that should be considered alongside several limitations. First, this study employed a mixed-method design and the results showed that the quantitative and qualitative data complemented each other nicely. A further strength of the study is that it recruited a relatively large sample. Nevertheless, in light of the potential pool of participants, the sample size was modest. Finally, this study was a good preliminary test of the Space ToC, providing important information about participants' perspectives of the outcomes gained from Space participation and helpful formative feedback that can already be actioned. Thus, the results certainly point to the need for further evaluation of Space with more stringent research designs.

The most important limitation that tempers the interpretation of these findings is the retrospective self-report design of this study. This means that the findings are subject to an increased risk for bias and confounding factors. Along these lines, there is an increased likelihood

that respondents to the survey were those participants who were either very satisfied or very dissatisfied with Space and, therefore, were more motivated to respond, compared to parents who were ambivalent about their experience with Space. Recruitment procedures limited participation to those participants with an email contact and internet access. Thus, participants without such technology were automatically excluded from participation. Finally, in terms of demographic characteristics; although the survey respondents were generally representative of Space participants overall, the sample was very homogenous in relation to New Zealand's ethnic diversity. Even though there were no significant differences in the results across sociodemographic characteristics, a more careful examination of how participants from Māori, Pacific, and other ethnic communities experience the programme is warranted. In parallel research, we found that both participants and facilitators recommended greater inclusion of te reo Māori, Māori cultural traditions/tikanga, and traditional caregiving practices in the Space curriculum (Amersfoort et al., 2021).

In conclusion, this study contributes to our understanding of the challenges first-time parents face, as well as the reasons why they engage in postnatal programmes. The results suggest that a supported playgroup programme has potential to provide both the social and educational context in which new parents can acquire caregiving knowledge and strategies, but also the necessary social support that can help them grow in parenting confidence and ease the transition to parenthood. However, as the findings suggest, facilitator skill and competency likely play an important role in achieving such positive outcomes for programme participants. Future research should focus on investigating this association more closely with more robust research designs and heterogenous samples to gain further evidence for the efficacy of the Space programme.

### **References**

- Amersfoort, L., Woods, A., & Friesen, M.D. (2021). *Community-based postnatal parenting support: Theory of change and case study evaluation of Space for You and Your Baby*. Unpublished Manuscript.
- Berlin, A., Törnkvist, L., & Barimani, M. (2016). Content and presentation of content in parental education groups in Sweden. *The Journal of Perinatal Education*, 25(2), 87-96. <http://doi.org/10.1891/1058-1243.25.2.87>
- Bohr, Y., Halpert, B., Chan, J., Lishak, V., & Brightling, L. (2010). Community-based parenting training: Do adapted evidence-based programmes improve parent-infant interactions? *Journal of Reproductive and Infant Psychology*, 28(1), 55-68. <http://doi.org/10.1080/02646830903294961>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77-101. <http://doi.org/10.1191/1478088706qp0630a>
- Centre on the Developing Child (n.d.) The IDEAS Impact Framework. Retrieved October 1, 2020, from <https://developingchild.harvard.edu/innovation-application/innovation-approach/components/>
- Commerford, J., & Robinson, E. (2016). Supported playgroups for parents and children: The evidence of their benefits. *Family Matters (Melbourne, Vic.)*, (99), 42-51.
- Copeland, D., & Harbaugh, B. (2019). "It's hard being a Mama": Validation of the Maternal Distress Concept in Becoming a Mother. *The Journal of Perinatal Education*, 28(1), 28-42. <http://doi.org/10.1891/1058-1243.28.1.28>
- De Haan, I. (2016). Supporting transition to parenthood in Aotearoa New Zealand. *Aotearoa New Zealand Social Work*, 28(3), 4-14. <http://doi.org/10.11157/anzswj-vol28iss3id243>
- De Silva, M.J., Breuer, E., Lee, L., Asher, L., Chowdhary, N., Lund, C., & Vikram, P. (2014). Theory of change: a theory-driven approach to enhance the Medical Research Council's framework for complex interventions. *Trials*, 15, 267-279. <http://doi.org/10.1186/1745-6215-15-267>
- Evans, S., Davies, S., Williams, M., Hutchings, J. (2015). Short-term benefits from the Incredible Years Parents and Babies Programme In Powys. *Community Practitioner*, 88(9), 46-48.
- Fonseca, A., Gorayeb, R., & Canavarro, M. C. (2015). Women's help-seeking behaviours for depressive symptoms during the perinatal period: Socio-demographic and clinical correlates and perceived barriers to seeking professional help. *Midwifery*, 31(12), 1177-1185. <https://doi.org/10.1016/j.midw.2015.09.002>
- Gardner, E., & Woolgar, M. (2018). Parenting in the community: A service evaluation of a universal, voluntary

- sector parenting intervention. *Journal of Community Psychology*, 46(3), 332-344. <http://doi.org/10.1002/jcop.21942>
- Glavin, K., Tveiten, S., Økland, T., & Hjälmhult, E. (2017). Maternity groups in the postpartum period at well child clinics – mothers' experiences. *Journal of Clinical Nursing*, 26, 3079-3087. <http://doi.org/10.1111/jocn.13654>
- Hanna, B.A., Edgecombe, G., Jackson, C.A., & Newman, S. (2002). The importance of first-time parent groups for new parents. *Nursing and Health Sciences*, 4, 209-214. <http://doi.org/10.1046/j.1441-0745.2002.00128.x>
- Hickey, A. (2019). The baby group: charting the typologies of a parental community of practice. *Leisure Studies*, 38(4), 480-493. <http://doi.org/10.1080/02614367.2019.1599045>
- Howard, L. M., & Khalifeh, H. (2020). Perinatal mental health: a review of progress and challenges. *World Psychiatry*, 19(3), 313-327. <https://doi.org/10.1002/wps.20769>
- Jackson, D. (2011). What's really going on? Parents' views of parent support in three Australian supported playgroups. *Australasian Journal of Early Childhood*, 36(4), 29-37. <http://doi.org/10.1177/183693911103600405>
- Jones, C., Erjavec, M., Viktor, S., Hutchings, J. (2016). Outcomes of a comparison study into a group-based infant parenting programme. *Journal of Child and Family Studies*, 25, 3309-3321. <http://doi.org/10.1007/s10826-016-0489-3>
- Kunseler, F.C., Willemen, A.M., Oosterman, M.I., & Schuengel, C. (2014). Changes in parenting self-efficacy and mood symptoms in the transition to parenthood: A bidirectional association. *Parenting: Science and Practice*, 14, 214-234. <http://doi.org/10.1080/15295192.2014.972758>
- Leahy Warren, P. (2005). First-time mothers: social support and confidence in infant care. *Journal of Advanced Nursing*, 50(5), 479-488. <http://doi.org/10.1111/j.1365-2648.2005.03425.x>
- Lee, K., Vasileiou, K., & Barnett, J. (2019). 'Lonely within the mother': An exploratory study of first-time mothers' experiences of loneliness. *Journal of Health Psychology*, 24(10), 1334-1344. <http://doi.org/10.1177/1359105317723451>
- McLean, K., Edwards, S., Evangelou, M., Skouteris, H., Harrison, L., Hemphill, S., ... Lambert, P. (2017). Playgroups as sites for parental education. *Journal of Early Childhood Research*, 15(3), 227-237. <http://doi.org/10.1177/1476718X15595753>
- McLean, K., Edwards, S., & Mantilla, A. (2020). A review of community playgroup participation. *Australasian Journal of Early Childhood*, 45(2), 155-169. <http://doi.org/10.1177/1836939120918484>
- Ministry of Education New Zealand. (n.d.). Establishing a certified playgroup. Retrieved October 1, 2020, from <https://www.education.govt.nz/early-childhood/running-a-service/starting-a-service/establishing-a-playgroup/the-value-of-playgroups/>
- Mize, J., & Pettit, G. (2010). The mother-child playgroup as socialisation context: A short-term longitudinal study of mother-child-peer relationship dynamics. *Early Child Development and Care*, 180(10), 1271-1284. <http://doi.org/10.1080/03004430902981470>
- Parfitt, Y., & Ayers, S. (2014). Transition to parenthood and mental health in first-time parents. *Infant Mental Health Journal*, 35(3), 263-273. <http://doi.org/10.1002/imhj.21443>
- Paris, R., & Dubus, N. (2005). Staying connected while nurturing an infant: A challenge of new motherhood. *Family Relations*, 54(1), 72-83. <http://doi.org/10.1111/j.0197-6664.2005.00007.x>
- Sanders, R., Lehmann, J., Gardner, F. (2014) Parents' experiences of early parenthood – preliminary findings. *Children Australia*, 39(3), 185-194. <http://doi.org/10.1017/cha.2014.20>
- Signal, T. L., Paine, S. J., Sweeney, B., Muller, D., Priston, M., Lee, K., ... Huthwaite, M. (2017). The prevalence of symptoms of depression and anxiety, and the level of life stress and worry in New Zealand Maori and non-Maori women in late pregnancy. *Australia New Zealand Journal of Psychiatry*, 51(2), 168-176. <http://doi.org/10.1177/0004867415622406>
- Spry, C., Morawska, A., & Sanders, M.R. (2010). *Participants Notes for Group Baby Triple P Provider training*. Milton, QLD: Triple P International Pty. Ltd.
- Statistics NZ. (2020, 1 September). *Skill levels of New Zealand jobs*. <https://www.stats.govt.nz/reports/skill-levels-of-new-zealand-jobs>
- Webster-Stratton, C. (2008). *The Incredibly Years Parents, Babies, and Toddlers series. Leader guide manual, 1st edn*. Seattle: Incredible Years.
- Wilkins, C., (2006). A qualitative study exploring the support needs of first-time mothers on their journey towards intuitive parenting. *Midwifery*, 22(2), 169-180. <http://doi.org/10.1016/j.midw.2005.07.001>
- Williams, K. E., Berthelsen, D., Viviani, M., & Nicholson, J. M. (2018). Facilitated Parent-child Groups as Family Support: A Systematic Literature Review of Supported Playgroup Studies. *Journal of Child and Family Studies*, 27(8), 2367-2383. <http://doi.org/10.1007/s10826-018-1084-6>
- Wright, A., Warren, J., Burriel, K., & Sinnott, L. (2019). Three variations on the Australian supported playgroup model. *International Journal of Social Welfare*, 28(3), 333-344. <http://doi.org/10.1111/ijsw.12361>

### Corresponding Author

Myron Friesen

Email: [myron.friesen@canterbury.ac.nz](mailto:myron.friesen@canterbury.ac.nz)

Phone: +6433695598



# Clarifying the Scope of Generalizability Theory for Multifaceted Assessment

Duncan J. R. Jackson<sup>1</sup>, George Michaelides<sup>2</sup>, Chris Dewberry<sup>3</sup>, and Paul Englert<sup>4</sup>

<sup>1</sup>King's Business School, King's College London

<sup>2</sup>Norwich Business School, University of East Anglia

<sup>3</sup>Independent Scholar

<sup>4</sup>Department of Psychology, Nanyang Technical University, Singapore

Generalizability theory (G theory) continues to be underutilized in applied psychological research, both in New Zealand and internationally, possibly due to uncertainties about the types of questions that it can be used to address. G theory and its associated random effects model basis is often positioned as an approach limited to the study of reliability. In contrast, latent variable theory, and its confirmatory factor analytic (CFA) basis, is used more widely to address issues of validity whilst controlling for reliability. This study clarifies the types of questions to which G theory can be applied by testing whether there is any justification for differences in interpretation between results based on G theory and latent variable theory. We reanalyzed data from an operational assessment center (N = 214 managerial assessesees) and found comparable aggregated effects, generalizability coefficients, and latent scores across the G theory and latent variable theory approaches, suggesting that both can be applied to problems related to reliability and structural validity.

**Keywords:** *Psychological assessment, multifaceted assessment, generalizability theory*

## INTRODUCTION

In applied psychology research and practice, the measurement of job-relevant characteristics is often complex and multifaceted. For example, assessment centers (ACs) utilize a complex design in which the ratings assigned to participants are a function of multiple interacting influences, such as raters, rating items, performance dimensions, and management simulation exercises (Lance, Foster, et al., 2007). Other relevant examples include personality assessments, job performance measures, situational judgment tests, and gamified assessments (Christian et al., 2010; Gnambs, 2015; Jackson, Kim, et al., 2016). Multifaceted assessments are widely applied in New Zealand as well as internationally (Krause & Thornton, 2009; Taylor et al., 2002). More general concerns about exercising statistical control over AC scores are relevant to the indigenous Māori population of New Zealand. Investigations into subgroup differences in this context have been explored in previous work on ACs (Jackson & Englert, 2011) and in other measures used in employee selection (Guenole et al., 2003).

The complex, multifaceted design of many organizational measurement systems presents a considerable challenge to those seeking to establish the extent to which they are valid and reliable. Generalizability theory (G theory) was originally developed specifically to address multifaceted measurement designs (Cronbach et al., 1972; Cronbach et al., 1963) and is therefore well-suited to such procedures commonly observed in organizations. Fairly recent developments around the application of G theory to ill-

structured measurement designs broaden its applicability, given how common these designs are in organizations (Putka et al., 2008). Yet, compared to applications of the more widely applied latent variable theory, G theory retains the status of the “underdog” with fewer research studies employing its use. As rough indication, a recent no-limits search of Business Source Complete with the keywords “generalizability theory” and “organization” only returned 92 hits. Replacing the former search term with “confirmatory factor analysis” increased the hit rate to 1,028.

In latent variable theory (e.g., Borsboom, 2008), CFA is routinely used to examine both reliability and validity. On the other hand, in G theory (e.g., Brennan, 2001), random effects models (REMs) are often used but, conceptually, their application is routinely restricted to an examination of reliability. In this article, we explain the relative advantages of G theory and REMs over latent variable theory and CFA in the psychometric evaluation of multifaceted measurement systems. We discuss a possible reason why G theory has been underutilized, particularly in examining issues relating to validity. One noteworthy explanation in this respect is a concern that the REMs utilized in analyzing G theory models may not produce results which are comparable to those generated with CFA. Directly addressing this issue, we examine the extent to which REMs and CFAs produce equivalent outcomes by reanalyzing a real-world data set using both approaches.

## Conceptions of Validity

The precise meaning of validity is complex and the focus of ongoing debate (Borsboom et al., 2004). Putka



and Sackett (2010, p. 39) define validity as “the degree to which evidence supports inferences one proposes to draw about the target of assessment”. Central to this definition is that the researcher is compelled to provide sufficient evidence to support the validity-related claims they make about their measurement procedure (Eignor, 2013).

Multiple forms of evidence might be used to support the case for the existence of hypothetical constructs (Campbell & Fiske, 1959; Cronbach & Meehl, 1955), including face, content, predictive, discriminant, and convergent validity elements. Another common form of validity evidence concerns the structure of ratings or responses within a given assessment procedure, akin to the concept of “substantive coherence” internal to the measure itself (e.g., Finch & French, 2015, p. 152). For example, the researcher’s focus might be on the extent to which ratings in an AC support the assumption that the raters are evaluating candidates on performance dimensions rather than on exercise performance (e.g., Lance et al., 2004). Such evidence facilitates an understanding about how measures function internally. An understanding about the internal structure of measures offers insights into why criterion-related relationships with external measures might be evident and so can be used effectively in conjunction with other forms of validity evidence (Putka & Sackett, 2010). Thus, while structure represents a single form of evidence, it might nevertheless be critical, particularly if the researcher investigates how the measurement structure interacts with other forms of validity evidence.

### **Conceptions of Reliability**

Reliability is traditionally defined as being concerned with measurement error, or variance that interferes with the assessment of constructs focal to the researcher’s aims (Borsboom & Mellenbergh, 2002; Schmidt et al., 2000). Putka and Sackett (2010) summarize contemporary, operational perspectives on reliability as relating to replication, expectation, and consistency. *Replication* refers to the reproducibility of an observation relating to a given construct. *Expectation* refers to the ability to infer from (a) observations (e.g., items, raters) used in a procedure to a hypothetical population of observations deemed as admissible for measuring a construct of interest, and (b) observations in a sample to those in a population of participants. *Consistency* refers to those elements of the measurement procedure that replicate and thus either contribute to construct measurement (i.e., an estimate of true score variance, see Spearman, 1907) or, less desirably, to some consistent but construct-irrelevant source of variation. Conversely, elements of the measurement procedure that fail to replicate contribute to undesirable error variance in observations.

In classical test theory and in G theory, reliability is represented by the ratio of true score (referred to as universe score in G theory) score to total variance (i.e., sources of universe score / sources of universe score + sources of error, see Crocker & Algina, 1986). This ratio is often referred to as a generalizability coefficient or G coefficient (Shavelson & Webb, 1991). In classical test theory, reliability is typically estimated for different purposes or perspectives on reliability using separate reliability coefficients. For example, coefficient alpha is

applied to questions about internal consistency, whereas test-retest reliability coefficients are applied to questions about temporal stability (Nunnally & Bernstein, 1994). In contrast, G theory allows the researcher to estimate and thus control for multiple perspectives on reliability simultaneously (e.g., in G theory it is possible to estimate effects relevant to internal consistency and temporal stability within the same analysis, Cronbach et al., 1972). This can present a more controlled perspective on reliability, particularly in complex, multifaceted measurement designs.

### **Applications of Latent Variable Theory and CFA**

Many of the measurement procedures used in applied psychology reflect a simple measurement design often involving items, constructs, and respondents. During the early and middle parts of the 20<sup>th</sup> century, classical test theory was applied to this type of design. The central assumption of classical test theory is that a person’s score on a test is a function of their true score on a latent construct or trait (e.g., conscientiousness) plus error. Here, error is viewed as being a consequence of multiple unmeasured variables associated with test administration, the candidate, and the test itself.

More recently, the development of CFA has made it possible to separate a general estimate of error into separate components, allowing a more detailed test of latent variable theory (Brown, 2006; Lance et al., 2002). By combining the error of measurement associated with each of the items involved in assessing individuals on a particular latent trait with residual error, an overall index of the reliability of the measure in evaluating that trait is obtained via CFA (Brown, 2006). Further, by examining how well relevant data sets fit the model proposed to measure the latent trait, CFA can be used to assess one form of validity evidence relating to the structure of the instrument (Borsboom & Mellenbergh, 2002).

CFA is of considerable utility in examining the reliability and validity of relatively simple measurement designs. However, its application can be limited in more complex measurement systems of the type often used in organizations. For example, in structured interviews (e.g., Saunders & Townsend, 2016), two or more raters may evaluate candidates against groups of items nested within several dimensions (e.g. communication skills, teamwork etc.). As the number of variables involved in a measurement design grows, so do the complexities involved in establishing the validity and reliability of that design. Here, the reliability of interviews depends on multiple, systematic measurement components, including raters, items, dimensions, all possible interactions between these elements, and residual error due to other unknown influences.

### **G Theory as an Approach to Reliability**

In the organizational literature, as well as in others, CFA is widely applied as an indication of construct-related evidence (e.g., Borsboom, 2008; Brown, 2006; Eid et al., 2008; Lance, Foster, et al., 2007; Lance, Woehr, et al., 2007). However, less clarity surrounds the purpose of G theory in addressing issues concerning reliability or validity. At its inception, G theory was primarily presented as a framework for understanding *reliability* in multifaceted measurement. Cronbach et al. (1963)

described G theory as a “liberalization of reliability theory” (p. 137), and primarily framed their arguments for the development of the theory in terms of reliability. As they developed G theory, Cronbach et al. (1972, p. 15) further positioned it as being concerned primarily with reliability, drawing attention to the flexible approach it provides, in that, based on judicious reasoning, theory, or research evidence, researchers can specify multiple sources of universe score (the G theory analogue of true score) and error. Classical test theory, on the other hand, usually offers no such flexibility (Brennan, 2000).

Other researchers and methodologists followed Cronbach et al. (1972) in presenting G theory as being principally concerned with the study of reliability. Brennan (2001) discusses the idea that conditions of measurement influence error or variability in scores, and that it is possible for researchers using G theory to quantify such influences. On summarizing the aims of G theory, Brennan states that “historically these types of issues have been subsumed under the heading “reliability”. Generalizability theory liberalizes and extends traditional notions of reliability” (p. 2). Similarly, and consistent with the Cronbach et al. description, Shavelson and Webb (1991) make reference to the focus in G theory on the dependability of scores. They state that the G coefficient often reported in G theory analyses is “analogous to classical test theory’s reliability coefficient” (p. 2).

In some of the most recent treatments of G theory in organizational contexts, researchers continue to frame the approach as a perspective on score reliability. Putka and Hoffman (2013, p. 115) separated measurement error in a G theory model into components classified as “reliable and unreliable”. Similarly, Putka and Hoffman (2014) framed their chapter on the application of G theory to job performance measures as a perspective on reliability. Akin to the perspective presented by Putka and colleagues, Jackson, Michaelides, et al. (2016) and Jackson et al. (2020) presented their G theory models as perspectives on reliable and unreliable sources of variance related to ACs and multisource performance ratings respectively, implying that G theory primarily concerns reliability. LoPilato et al. (2015, p. 693) defined G theory as a “statistical framework for identifying factors that affect the reliability of measurements”. Woehr et al. (2012) stated that “Typically, G-theory is introduced and discussed in the context of reliability estimation” (p. 15).

### **G theory as an Approach to Validity**

Although, the descriptions offered above suggest that G theory is primarily concerned with reliability, not all researchers describe the approach as being restricted to the reliability domain, and indeed several scholars position it primarily as an approach towards summarizing validity evidence. Arthur et al. (2000, p. 819) had as one of their

research objectives “to recommend and demonstrate the use of generalizability theory analysis to assess convergent/discriminant validity” in the context of AC ratings. They expanded on this description, noting that “Evidence of construct-related validity is derived from the extent to which variance associated with the constructs of interest (measurement focus) is large relative to the variance associated with conditions of measurement<sup>1</sup>”. Lievens (2001b, p. 203) aimed to “shed light on the issue of assessment center construct validity” using G theory as a basis. Similarly, Lievens (2001a) applied G theory, in part, to examine evidence of “discriminant validity” in ratings from assessor training (p. 259). In the context of multitrait-multimethod matrices (MTMMs), Woehr et al. (2012) investigated the question: “How do the variance components stemming from G-theory relate to the traditional notions of construct-related validity?” (p. 141), and demonstrated how effects estimated via G theory have analogs in classic work on MTMMs (e.g., Campbell & Fiske, 1959). Highhouse et al. (2009) described G theory as “an especially powerful method for gathering construct validity evidence” (p. 784).

### **The Applicability of G Theory**

While we discuss reliability and validity separately above, this distinction is not altogether clear in the psychometric literature, with Campbell and Fiske (1959) describing it in terms of “regions on a continuum” (p. 83). This idea is reflected in the developmental stages of G theory, where Cronbach et al. (1963) noted that “the theory of ‘reliability’ and the theory of ‘validity’ coalesce” in the context of G theory<sup>2</sup> (p. 157), and Cronbach et al. (1972) stated that “generalizability theory blurs the distinction between reliability and validity” (p. 380). An elaboration of this latter statement was offered by Brennan (2000). In a typical G theory-based analysis, multiple, systematic facets<sup>3</sup> are isolated in a data set. Brennan suggests that some of these facets might be associated with validity (e.g., Participant  $\times$  Trait interactions) and others with reliability (e.g., Participant  $\times$  Item interactions).

Notwithstanding these observations, recent and historical perspectives on G theory suggest that the approach is primarily concerned with reliability (e.g., Cronbach et al., 1972; Cronbach et al., 1963; Jackson et al., 2020; Jackson, Michaelides, et al., 2016; LoPilato et al., 2015; Putka & Hoffman, 2013, 2014; Putka & Sackett, 2010; Thompson, 2003). This perhaps limits its perceived usefulness. Therefore, and only for the purposes of comparison in this paper, we begin by assuming the popular perspective that the purpose of G theory is to summarize reliability evidence. In Table 1, we present effects relevant to an example task-based AC model (Jackson et al., 2010; Thoresen & Thoresen, 2012) and

<sup>1</sup> We infer here that “conditions of measurement” refer to those measurement conditions not specified as relating to constructs of interest (e.g., variance related to items, raters, etc.).

<sup>2</sup> Here, Cronbach et al. (1963, p. 157) specifically refer to the idea that the universe of admissible observations is a construct domain introduced by the researcher that has

potential “explanatory or predictive power”. A G theory analysis therefore offers suggestions about “how validly one can interpret a measure as representative of a certain set of possible measures” (see p. 157).

<sup>3</sup> A *facet* is any systematic source other than participants that contributes to variance in scores (e.g., items, raters, etc.).

**Table 1.** Generalizability Theory and Confirmatory Factor Analytic Perspectives on Task-Based Assessment Center Effects

Effect	Common interpretation	G theory perspective, often associated with <sup>a</sup> :	CFA perspective, often associated with:
p	General factor	Reliability (relative <sup>b</sup> )	Validity, structural
pc	Role-exercise-dependent interaction	Reliability (relative)	Validity, structural
pi:c,e	General, item, and role-exercise interaction + residual variance	Reliability (relative)	Reliability
c	Role-exercise main effect	Reliability (absolute <sup>c</sup> )	NA
i:c	Item-in-role-exercise main effect	Reliability (absolute)	NA

*Note.* In task-based assessment centers, exercise factors represent role constructs of interest and are not considered to be method or mode effects. p = participant; c = role-exercise construct; i = rating item; e = residual error variance; G theory = generalizability theory; CFA = confirmatory factor analysis; NA = non-applicable. Exercise and item main effect estimates are unavailable in typical CFA output. <sup>a</sup>We acknowledge that there is considerable variability in the literature here, with several authors positioning G theory as capable of summarizing validity evidence. We refer here to some of the original (e.g., Cronbach et al., 1963) and the most recent (e.g., Putka & Hoffman, 2013) perspectives on this issue. <sup>b</sup>Variance components associated with relative decisions apply where the aim is to evaluate the score of an individual in terms of how it relates to scores from a larger group (e.g., norm-referenced scores). <sup>c</sup>Absolute decisions, on the other hand, are concerned with cut-off scores (e.g., pass/fail criteria), which are arguably less common in studies of organizations (see Shavelson & Webb, 1991 for further discussion on relative versus absolute decisions).

compare the hypothetical interpretation of these effects from a reliability-oriented G theory perspective against a more widely applied latent variable theory perspective analog. Of the three effects that are available for comparison across the two perspectives, only one, that for residual error, shares the same interpretation across the G theory and latent variable methodological frameworks.

Assuming that the reliability and validity concepts are meaningfully distinguished from one another, cross-theory differences in the interpretation of effects raises a conundrum. We suggest that evidence for reliability should be interpreted according to an accepted definition of reliability, regardless of the approach used to garner that evidence. Likewise, evidence for validity should be interpreted as it relates to an accepted definition of validity, and the status of such evidence should not depend on the approach used in its collection. Variability in the interpretation of effects in this respect could impede progress in understanding organizational phenomena.

**Comparing G theory- and Latent Variable Theory-Related Methods**

Why is it that output from methods associated with traditional and recent perspectives on G theory is framed as an examination of reliability (e.g., LoPilato et al., 2015), whereas output from methods associated with latent variable theory is often interpreted as it relates to an examination of validity (e.g., Borsboom, 2008)? It is possible that that the REMs popularly applied in G theory versus the CFAs in latent variable theory simply produce fundamentally different results. Output from these methods could lend itself more towards an interpretation based in reliability in G theory, and validity in latent variable theory.

Several researchers have replicated results from REMs using constrained CFA models (Marcoulides, 1996; Raykov & Marcoulides, 2006; Woehr et al., 2012). Notably, in the context of MTMMs used in organizations, Woehr et al. replicated the variance estimates in a univariate<sup>4</sup> REM model with a constrained CFA model. Thus, the capacity for CFA to reproduce REM results is known. However, the ability to reproduce the same variance estimates across REM and CFA addresses only a

component of the problem discussed here. Two key issues here are how those results are interpreted (i.e., as reliability and/or validity evidence); and whether there is any justification for interpreting results differently based on the method from which they have been derived, and the specific theoretical framework on which a given method is based.

A relevant consideration is that in G theory, aggregation formulae are often applied to REM variance estimates in a manner that is not typical or even clearly possible in a traditional latent variable theory framework via CFA. Aggregation can greatly influence relative effect size in a measurement model (Kuncel & Sackett, 2014; Putka & Hoffman, 2013). The effect estimates in a REM are orthogonal and this statistical property enables aggregation formulae to be selectively applied to relevant effects (Brennan, 1992, 2001; Searle et al., 2006). In principle, it is possible to apply G theory-based aggregation formulae to estimators generated via a CFA constrained in a manner analogous to a corresponding REM, and then to compare outcomes from both types of analysis. It would be possible here to establish whether there is any justification for interpreting effects differently across methods, given the application of formulae usually applied in G theory.

An issue related to aggregation formulae in G theory centers on G coefficients. The G coefficient is widely applied to analyses invoking the G theory framework (Brennan, 2001). Whether applying G coefficients based on REM versus CFA estimators makes a difference to statistical outcomes is currently unclear. If, overall, REM and constrained CFA results are similar, the justification for interpreting one type of analysis differently from another, depending on whether the researcher takes a G theory or latent variable theory perspective, is weakened.

Both REMs and CFAs are used to indicate variance associated with constructs in G theory and latent variable theory respectively (Borsboom, 2008; Cronbach et al., 1972). It is possible to generate latent scores (sometimes referred to as factor scores) for these construct effects both in REMs and in CFAs. Latent scores are defined as an estimate of a participant’s relative standing on a construct

<sup>4</sup> We focus on univariate REMs, given the similarities between multivariate REMs and their widely-criticized

correlated uniqueness CFA analog (Lance et al., 2002; Woehr et al., 2012).

of interest. In conceptual terms, latent scores provide an indication of what a participant's score would have been on the construct of interest, had it been possible to measure it directly (Brown, 2006) and are relevant to constructs evaluated via multifaceted assessment. A consideration of latent scores in REMs and G theory is rare (however, see Ward, 1986) and we were unable to find any sources where REM- and CFA-derived latent scores had been compared. Such a comparison could shed light on what is perhaps the core purpose of many multifaceted measures: their capacity to produce intended construct scores. Differences in effect size and patterns of intercorrelation between latent scores generated through REM versus CFA might offer suggestions about the basis for differences in the interpretation of their respective outputs. This could, in turn, highlight whether there are fundamental differences between REMs and CFAs that justify restrictions in the scope of application associated with G theory.

### Summary

Under a latent variable theory perspective, CFA is regularly considered to be concerned with structural validity as well as reliability (Borsboom, 2008; Borsboom et al., 2004; Eid et al., 2008; Kleinmann & Köller, 1997; Lance, Woehr, et al., 2007). Historical and recent perspectives on G theory, via the interpretation of REMs, position it primarily as a perspective on reliability (Thompson, 2003). The latter perspective restricts the scope of G theory relative to latent trait theory, in terms of the types of research questions that it can address. This might have limited the popularity of G theory, despite the fact that the REMs commonly used by G theorists are, in many circumstances, more accommodating of the complex research designs often encountered in organizational research (Michalak et al., 2019; Soltani et al., 2005). If a comparison between estimators generated using a REM and CFA reveals little difference in outcomes, even when considering aggregation, G coefficients, and latent scores, then this would call into question differences in interpretation from G theory versus latent variable theory standpoints. In keeping with these arguments, we propose the following, three Research Questions (RQs):

*RQ1:* When comparing aggregated results across REM, constrained CFA, as well as traditional CFA output, is there any justification for interpreting effects differently across methods as they relate to reliability or validity?

In RQ1, as we expand on below, our intention is to create two analyses: one based on a REM and the other based on a CFA, that constrain their estimates in a similar manner. The intention is to create variance component estimates that are directly comparable, but that have been generated using different estimation processes.

*RQ2:* Do G coefficients based on REM and CFA return similar outcomes?

In RQ2, our aim is to use the variance components mentioned for RQ1 to generate G coefficients that are directly comparable across estimates based on REM and CFA.

*RQ3:* Do latent scores based on REM and CFA generate similar effects and patterns of intercorrelation?

In RQ3, we aim to produce latent scores that are directly comparable for analyses based on REM and CFA so that they can be contrasted against one another and correlated.

### METHOD

Our data-analytic aims in this study center on providing a comparison between effects generated using REMs and effects generated using CFA. For this purpose, we reanalyzed a subset of data from Jackson et al. (2010). Our interest here was in testing a model with a small number of effects so that it could be easily reproduced in both REMs and CFAs and to maintain simplicity and brevity. In the original study, the authors analyzed data from a task-based AC, which is a simplified version of a traditional AC, where role constructs that are assessed within each exercise. Thus, scores for each exercise in a task-based AC represent role-exercise constructs (Jackson, 2012). We provide a brief description of participants and materials below. A full description of the AC under scrutiny is available in the Jackson et al. (2010) article. We note here that our aims are not oriented towards contributing to the literature on the structural characteristics of ACs and our inclusion of data related to a task-based AC is incidental.

### Participants

A total of 214 managerial assesseees from New Zealand participated in the study (we removed ratings from 1 participant due to incomplete data, bringing our analysis  $N$  to 213). The mean age of participants was 45.53 ( $SD = 10.33$ ) and 54% of the sample were men, 46% were women. The organization under scrutiny specialized in postal, insurance, credit, banking, and administrative services. Assesseees were evaluated by 19 assessors ranked one level above assesseees and 4 additional assessors who were employed as consultant psychologists. We could not estimate assessor-related effects because the ratio of assessors to assesseees was set at 1:2 to reduce costs for the participant organization. However, recent research across multiple samples suggests that assessor-related effects tend to be small (see Jackson, Michaelides, et al., 2016; Putka & Hoffman, 2013), assuming that assessors are adequately and appropriately trained. In the present case, assessors were trained using a frame-of-reference training (FORT) procedure, as recommended in the assessor training literature (Gorman & Rentsch, 2009; Pulakos, 1986). Training lasted for a 2-day period and covered familiarization with assessment materials, common rater errors, and mock assessments with related FORT discussions.

### AC Characteristics

AC ratings related to (a) a group discussion and oral presentation based on managing new staff (i.e., the management role), (b) a group discussion and oral presentation on selecting new staff (i.e., the human resource selection role), and (c) a group-based problem-solving exercise (i.e., the contextualized problem-solver role). Thus, the role-exercises included 3 levels

represented for each exercise. For each role-exercise construct, 7 behavioral descriptor items (21 items in total, e.g., *uses objective and non-emotive language when delivering feedback to others*) were retained for analysis. Behavioral descriptors were rated on a scale ranging from 1 (certainly below standard) to 10 (certainly above standard). All exercises were developed based on competency and inductive job analyses (Tett et al., 2000; Williams & Crafts, 1997).

**Analyses**

Our primary interest was in comparing two analogous models: one based on REMs, the other on CFA. The first model comprised a REM (see Searle et al., 2006) with restricted maximum likelihood (REML) estimation as a representation of the models typically used in contemporary studies using G theory (e.g., Putka & Hoffman, 2013, 2014). A total of 3 main effects were estimated in this model, relating to participant assesseses (*p*), role-exercise constructs (*c*), and rating items nested in role-exercise constructs (*i:c*)<sup>5</sup>. Taking interactions between effects into account, this resulted in a total of 5 effects that could be estimated within the REM model, each of which is listed and described in Table 1.

The second model that we tested was based on a CFA constrained to enable estimation in a manner analogous to that relevant to the REM (Marcoulides, 1996; Raykov & Marcoulides, 2006). This involved constraining the CFA model to have equal latent factor variances and unique variances. All factor covariances and error covariances were constrained to zero and all factor loadings were constrained to 1 (see Woehr et al., 2012, p. 144, Figure 2 caption). It was possible to estimate 3 effects with this approach, including the analogs of the main effect for *p*, the *pc* interaction, and an estimate for residual variance (see Table 1 for a description of these effects). To add a supplementary perspective, we tested a regular CFA model with correlated latent factors (as depicted in Figure 1).

To the REM and CFA variance estimates, we applied aggregation and G coefficient formulae based on those from the extant G theory literature (Brennan, 2001; Jackson et al., 2020; Putka & Hoffman, 2013; Shavelson & Webb, 1991). We extracted latent scores relating to role-exercise constructs from both the REM and the constrained CFA and correlated the two sets of latent scores. For the REM analysis, latent scores were derived from random intercepts relating to Participant × Exercise interactions (e.g., Liu et al., 2008). For CFA, latent scores were represented for each construct by the average of the product of each item response and its associated factor loading (e.g., Brown, 2006). The REM was conducted using the lmer function in lme4 for R (Bates et al., 2015). The CFA was conducted using lavaan for R (Rosseel, 2012). G coefficients were specified such that the effects for *p* and *pc* defined universe score. This is because *p* represents general individual differences, which is routinely of focal interest in an evaluation approach (Shavelson & Webb, 1991). The *pc* interaction represents individual differences on the focal constructs of interest, and thus represents a source of value to the evaluation instrument (Putka & Hoffman, 2013). The residual effect was specified as contributing to error.

**RESULTS**

To provide a perspective on goodness-of-fit, we tested the model shown in Figure 1, which represents the standard CFA model implied in the task-based AC literature with correlated, latent role-exercise constructs (e.g., Jackson et al., 2010; Thoresen & Thoresen, 2012). The model converged within expected parameters and model fit was acceptable according to criteria specified in Brown (2006)<sup>6</sup>,  $\chi^2 = 245.74(165)$ ,  $p < .001$ ; comparative fit index (CFI) = .970; Tucker-Lewis index (TLI) = .962; root-mean-square error of approximation (RMSEA) = .045; standardized root-mean-square residual (SRMR) = .043. Averaged, squared standardized loadings suggested effect sizes for the general factor = .19, role-exercise constructs = .40, and unique variance = .41. Averaged,

**Table 2.** Comparison of Generalizability Theory and Confirmatory Factor Analytic Effects

Effect	Generalizability theory estimates					Constrained CFA estimates			
	VC	Total %	BP %	Aggregation formula	BP % Aggregated	VC	BP %	Aggregation formula	BP % Aggregated
$\sigma_p^2$	.4224	26.08	27.71	$\sigma_p^2$	45.11	.4200	27.69	$\sigma_p^2$	45.09
$\sigma_{pc}^2$	.4159	25.68	27.28	$\sigma_{pc}^2$	44.42	.4140	27.29	$\sigma_{pc}^2$	44.44
$\sigma_{pi:c,e}^2$	.6862	42.37	45.01	$\sigma_{pi:c,e}^2/n_{i:c}$	10.47	.6830	45.02	$\sigma_{pi:c,e}^2/n_{i:c}$	10.47
$\sigma_c^2$	.0216	1.33	–	–	–	–	–	–	–
$\sigma_{i:c}^2$	.0736	4.55	–	–	–	–	–	–	–
G					.90				.90

Note. *p* = participant main effect (or general performance effect); *c* = exercise-role construct; *i* = item; *e* = residual error; G = generalizability coefficient; VC = variance component; BP = between-participant variance. Dashes indicate non-applicability. Generalizability theory estimates are derived from the variances in a random effects model. Confirmatory factor analysis (CFA) estimates are derived from variances in a CFA model constrained as described below. The residual in the CFA analysis is estimated using the formula  $1 - \sigma_p^2 - \sigma_{pc}^2$ . Constraints imposed on the CFA model included equal *c* factor variances, equal unique variances, *c* factor covariances constrained to zero, *c* factor and *p* factor covariances constrained to zero, all error covariances constrained to zero, and all factor loadings constrained to one.

<sup>5</sup> In G theory notation, the presence of a colon (:) indicates a level of nesting. For example *i:c* implies that items are nested in constructs.

<sup>6</sup> CFA-related goodness-of-fit was not estimated for the constrained models that follow because associated fit

indices can “reflect types of misfit that have little or no bearing on the accuracy of G-theory model parameter estimates” (Woehr et al., 2012, p. 158).



**Table 3. Factor Score Correlations**

REM latent scores			
	c1	c2	c3
c1			
c2	.61		
c3	.57	.42	
CFA latent scores			
	c1	c2	c3
c1			
c2	.61		
c3	.58	.48	
REM with CFA latent scores			
	c1	c2	c3
c1	.99	.59	.57
c2	.60	.99	.45
c3	.57	.45	.99

*Note.* c = role-exercise construct, G theory = generalizability theory, REM = variance components analysis, typically used in G theory; CFA = confirmatory factor analysis. In the bottom matrix, REM latent scores appear on the vertical axis. REM estimates based on latent scores for  $\sigma_p^2 + \sigma_{pc}^2$ , where p = participant. CFA latent scores were estimated from a correlated 3c model.

squared standard covariances among role-exercise constructs = .08.

The models used in REMs, and often as a basis for G theory, offer a somewhat different perspective on observed data than that associated with CFA. To allow for comparison between the CFA and REM analyses, we constrained the CFA model in Figure 1 as described in the note in Table 2, in keeping with guidance provided in the methodological literature (Marcoulides, 1996; Raykov & Marcoulides, 2006; Woehr et al., 2012). Table 2 shows a comparison between variance components from on a REML-based REM and variance components from an analogous, restricted CFA model. Both models converged acceptably.

We applied formulae to REM and analogous CFA estimates in Table 2 based on those commonly applied in the G theory literature (see Brennan, 2001; Cronbach et al., 1972; Shavelson & Webb, 1991), so as to approximate the effects of aggregation on variance estimates. In the present case, only aggregation to role-exercise scores was considered, because this is of focal interest in task-based ACs (Jackson et al., 2005; Lance, 2012). A total of 5 effects were available for the REM, which included 2 main effects that were not relevant to between-participant comparisons. The remaining 3 effects were relevant to between-participant comparisons and were available in both the REM and CFA analyses.

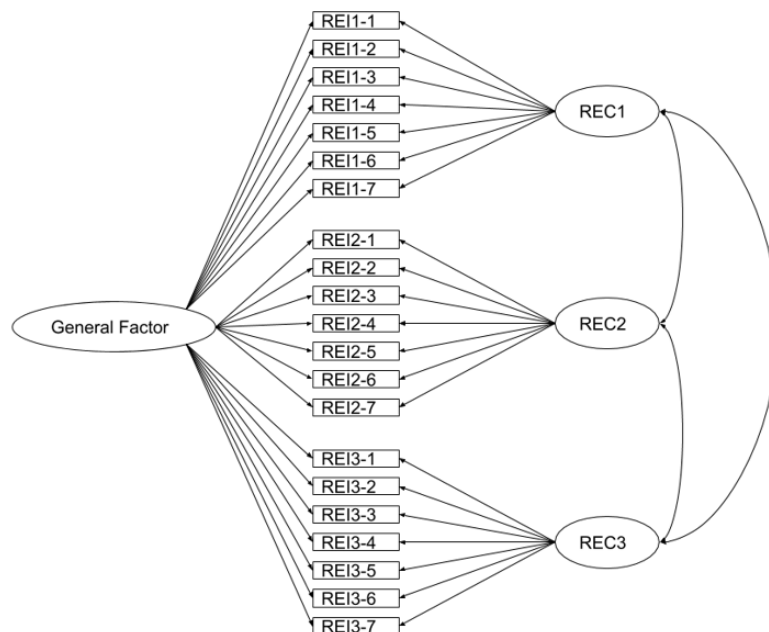
The results presented in Table 2 show outcomes that are almost identical when comparing across the G theory and CFA analyses. With respect to effect size, both pre-

and post-aggregated results only differed by a maximum of .02 of a percentage point, thus indicating near zero differences between outcomes generated by the two analytic approaches (in response to RQ 1). As expected for the task-based approach to ACs, the majority of variance on aggregation was associated with general performance (around 45%, regardless of estimation approach) and Participant  $\times$  Role-Exercise Construct interactions (around 44%, again regardless of estimation approach). Table 2 shows that formulae for aggregation commonly applied in the G theory literature can be applied in the same way to constrained CFA variance components with practically the same outcomes. It follows that G coefficients, estimated on both REM and CFA variance components (see RQ 2), in both cases, analogs of  $(\sigma_p^2 + \sigma_{pc}^2)/[\sigma_p^2 + \sigma_{pc}^2 + (\sigma_{pi:c,e}^2/n_{i:c})]$ , resulted in identical outcomes to 2dp at .90 (where p = participant, c = exercise-role construct, i = item, and e = residual error).

The results shown in Table 2 and the results of the traditional CFA model (shown in Figure 1) also reflect similar outcomes. However, the different methods need to be considered with respect to their treatment of data. In the REMs that act as the basis for G theory (shown in Table 2), it is assumed that any latent constructs under consideration do not share any common variance beyond that which is already accounted for by the general effect,  $\sigma_p^2$  (Marcoulides, 1990; Raykov & Marcoulides, 2006). In the case of the traditional CFA depicted in Figure 1, the effect size for role-exercise construct loadings = 39.63% and for the general factor = 19.38% (based on average, squared standardized loadings). The summary role-exercise effect here was of a greater magnitude than that presented in Table 2 for the G theory model (see results prior to aggregation: role-exercise effect = 27.71%, general factor = 27.28%). This is because the traditional CFA estimate for the role-exercise effect includes variance shared between role-exercise constructs. Once these method-specific idiosyncrasies are acknowledged, even the results of the traditional CFA are similar to those presented in Table 2 for the G theory analyses given that larger role-exercise effects are expected from a traditional CFA.

To provide an additional perspective on the outcomes above, we extracted latent scores for role-exercise constructs based on both REM and CFA estimates (see RQ 3). Table 3 shows three matrices, which display correlations between (a) REM latent scores, (b) CFA latent scores, and (c) REM latent scores and CFA latent scores. When comparing the separate REM and CFA outcomes (i.e., a and b above), it is clear in Table 3 that the two modes of estimation make very little difference to how the latent scores intercorrelate. The largest of these differences was between role-exercise constructs 2 and 3 ( $r = .42$  versus  $r = .48$ ). When expressed in terms of a percentage of variance explained, this is a near-zero difference (i.e.,  $< .004\%$ ). REM and CFA latent scores (i.e., c above), shown in the diagonal of the bottom matrix in Table 3, correlated at a uniform .99 for all 3 role-exercise constructs. This provides further evidence that the results across G theory and analogous CFA methods are, for practical purposes, almost identical.

**Figure 1.** Task-based assessment center confirmatory factor analysis model, showing role-exercise indicators (REI1-1 through REI3-7), role-exercise latent constructs (REC1 through REC3), and a general factor.



**DISCUSSION**

G theory has never reached the status of a mainstream methodological approach in applied psychology, despite a lengthy history and wide applicability to the complex measurement designs routinely found in organizations (Cronbach et al., 1972; DeShon, 2002; Putka & Hoffman, 2014). We posit that a key reason for this lack of uptake is because of uncertainties about what types of research questions G theory can be used to address. Both historically (Cronbach et al., 1963), and in recent organizational research (Jackson, Michaelides, et al., 2016; Putka & Hoffman, 2013, 2014), G theory has been characterized as an approach towards summarizing reliability evidence. However, some researchers position the approach as being relevant to summarizing validity evidence (Arthur et al., 2000; Highhouse et al., 2009; Lievens, 2001a, 2001b; Woehr et al., 2012). In contrast to the differing perspectives on the purpose of G theory, much more agreement is apparent about the role of CFA and its capacity to summarize structural validity-related evidence whilst also acknowledging reliability (e.g., Brown, 2006). It might therefore be no coincidence that CFA is more widely applied in the discipline (e.g., Lance et al., 2004; Lance et al., 2002) than is G theory (e.g., Murphy & DeShon, 2000).

We compared results from a G theory model based on a REM of a task-based AC (Jackson et al., 2010) with analogous results generated through a CFA model constrained to match the outcomes generated through the REM. Comparison of the REM and CFA outcomes, including those relating to aggregation formulae often applied in G theory (RQ 1), G coefficients (RQ 2), and latent scores (RQ 3), revealed that the two methods provided practically identical results (see Tables 2 and 3). We found that a regular CFA model with correlated latent factors suggested conclusions similar to those based on the REM.

Our results suggest that REM, the technique normally adopted when G theory is applied, provides a perspective that is analogous to that provided by CFA, and that there is, therefore, no cogent justification for cross-method differences in the interpretation of specific effects. Cronbach et al. (1972) stated that G theory blurs the reliability-validity distinction. Brennan (2000) suggested that Cronbach et al. referred here to the idea that G theory can address (a) sources of variance often considered to be about validity and (b) sources of variance often considered to be about reliability. Our results are consistent with Brennan’s interpretation,

and we offer the extension that irrespective of whether a G theory or CFA approach is used, any sources of variance related to observations (e.g., items, assessors) are likely to concern reliability, whereas any sources of variance related to the equivalent of latent constructs (e.g., dimensions, personality constructs, role-exercise constructs) are likely to concern structural validity.

In a G theory model, distinctions between sources of variance as they relate to validity or reliability might be straightforward in many cases because each effect is presented separately and can, potentially, be meaningfully categorized. For example, with reference to the between-participant effects listed in Table 2, the effects  $\sigma_p^2$  and  $\sigma_{pc}^2$  are concerned with the equivalent of CFA latent constructs and thus could be categorized as relating to validity evidence. The former of these effects represents the CFA analog of a general performance effect or *positive manifold* (e.g., Ree et al., 2015). The latter interaction represents the CFA equivalent of role-exercise latent constructs (Jackson, 2012). In contrast, the  $\sigma_{pi:c,e}^2$  effect includes the influence of indicator items, and it could therefore be argued that this effect relates to reliability evidence.

What is less clear, perhaps, is how G coefficients should be conceptualized. If we accept the classification of effects as sources of either reliability or validity evidence as described above, then G coefficients combine aspects of both reliability and validity. That said, there is often a predictable pattern to how G coefficients are constructed in that validity-related effects commonly define the numerator and reliability-related effects commonly define the denominator in G coefficient equations. This is certainly the case in the present example where the G coefficient  $(\sigma_p^2 + \sigma_{pc}^2) / [\sigma_p^2 + \sigma_{pc}^2 + (\sigma_{pi:c,e}^2 / n_{i,c})]$  contains validity-related effects in the numerator and the reliability-related effect in the

denominator<sup>7</sup>. Thus, one interpretation of the G coefficient could be the ratio of structural validity to reliability evidence.

The finding of a relatively large proportion of variance associated with what is presumed to be a latent construct does not guarantee, in any way, the validity of the measure being applied (Putka & Sackett, 2010). It does suggest a systematic source of variance that is potentially relevant to the internal structure of the assessment procedure, which, we suggest, could count as one, limited, source of validity evidence. The possibility still exists, however, that this systematic source of variance might, in fact, be irrelevant to the construct(s) of interest. Other sources of evidence will be necessary to determine the nature of such effects, whether they relate to what was intended for measurement, and whether they relate meaningfully and as expected to externally measured constructs (see Strauss & Smith, 2009).

### Implications

Our results suggest that G theory and CFA deal with sources of evidence for both reliability *and* structural validity. In future research involving G theory, researchers using either methodological approach could classify effects as they pertain to reliability or validity evidence, to assist in developing a clear and consistent understanding of the structure of multifaceted measures that does not depend on methodological context.

Our findings highlight the idea that the theoretical principles of G theory apply with the use of methods such as CFA, just as much as they apply when using REM. REM appears to have become synonymous with G theory, but, in fact, G theory is not REM. The “statistical machinery” (Brennan, 1997, p. 15) used to generate effects in G theory is secondary to the theory itself. As suggested in this paper, at least some G theory models can be adequately estimated using CFA. There are likely other statistical methods that could be used as a basis for G theory. Even within REM, there are different options that researchers can choose from to estimate effects, including those based on REML, ANOVA-analogous, or Bayesian estimators (Brennan, 2001). The main issue here, though, is that G theory should be thought of as a theoretical framework that is not anchored to a specific statistical method. While REM represents the most common basis for G theory, its aggregation formulae, G coefficients, and latent scores can be used with other statistical foundations, as we demonstrate with CFA.

Our results suggest that consideration should be given to the advantages and disadvantages of using one statistical basis over another for G theory. The benefits of employing CFA include that it can provide multiple perspectives on a data set, including a model constrained such that it is similar to a REM as well as a regular CFA model with correlated latent constructs. The latter model can provide more detail than REMs about each specific construct under scrutiny, as well as GFIs for the model as a whole (Le et al., 2009; Woehr et al., 2012). However, particularly with studies involving large numbers of

effects, REMs might present a more practical approach than CFA because fewer parameters require estimation in REMs. Moreover, organizational measurement often requires the use of raters (e.g., in job performance evaluation or ACs). The presence of multiple raters might present a measurement design that is ill-structured (i.e., neither perfectly crossed nor nested, see Putka et al., 2011; Putka et al., 2008). While REML or Bayesian estimators in REM can handle ill-structured designs, there is often no practical way to address such designs in CFA (Putka et al., 2011; Putka et al., 2008).

### Limitations

A limitation of our study is the simplicity of the model used to demonstrate comparisons between REM and CFA. However, we purposely chose a simple model (i.e., a model with a small number of effects) to facilitate an explanation of G theory, which is often described as conceptually complex (DeShon, 2002). Moreover, a small number of effects allows for direct comparisons between REM and CFA models, where such comparisons might not be practical with models that contain many effects. For example, it can be impractical to estimate effects related to raters with CFA because doing so could require a latent variable for each of potentially large numbers of raters (Jackson et al., 2020). The downside to the application of a simple model, however, is that we are unable to show from this study how different combinations of effects might contribute to universe score and error variance. Nonetheless, we are confident that the reader will be able to extrapolate in principle from the basic design presented here to more complex designs used in other operational assessment procedures.

For our G theory model, we could have explored alternatives to the REML estimators that we used. For example, Bayesian estimators have been recommended for more complex designs in the AC literature (Jackson, Michaelides, et al., 2016) and in the literature on multisource performance ratings (Jackson et al., 2020). Bayesian approaches provide an effective approach towards defining variability around effect estimates in the form of credible intervals that relate to a full posterior distribution (Gelman, 2006). However, empirical evidence suggests that G theory analyses based on Bayesian or REML estimators provide results that are similar or identical, assuming that none of the effects are fenced at zero (Jackson, Michaelides, et al., 2016; LoPilato et al., 2015). No fenced estimates were present in our analyses.

### Conclusion

G theory is underutilized in applied psychology. We see this as an oversight because it is well suited to many of the measurement designs encountered in organizations, both in New Zealand and internationally. G theory could therefore help inform on theory and practice in organizational measurement. More clarity is needed on the types of research questions that G theory can be used to address, albeit those concerning reliability and/or validity evidence. Our results suggest that G theory can

<sup>7</sup> Note that the object of measurement here is participants (p) and at least some effects relating to p almost always define universe score. In G theory, it is possible to combine

different elements of universe score and error, but the onus is on the researcher to justify this classification.

be used to evaluate both reliability and structural validity evidence in a similar manner to how CFA is routinely applied. Effects representing observations can be categorized as relating to reliability and effects

representing analogs of latent constructs can be categorized as relating to one type of structural validity evidence, similar to the latent constructs addressed by CFA (see Strauss & Smith, 2009).

## References

- Arthur, W., Jr., Woehr, D. J., & Maldegan, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical reexamination of the assessment center construct-related validity paradox. *Journal of Management*, 26(4), 813-835. [https://doi.org/10.1016/S0149-2063\(00\)00057-X](https://doi.org/10.1016/S0149-2063(00)00057-X)
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives*, 6, 25-53. <https://doi.org/10.1080/15366360802035497>
- Borsboom, D., & Mellenbergh, G. J. (2002). True scores, latent variables and constructs: A comment on Schmidt and Hunter. *Intelligence*, 30(6), 505-514. [https://doi.org/10.1016/S0160-2896\(02\)00082-X](https://doi.org/10.1016/S0160-2896(02)00082-X)
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity [Psychometrics & Statistics & Methodology 2200]. *Psychological Review*, 111(4), 1061-1071. <https://doi.org/10.1037/0033-295X.111.4.1061>
- Brennan, R. L. (1992). *Elements of generalizability theory*. American College Testing (ACT) Publications.
- Brennan, R. L. (1997). A perspective on the history of generalizability theory. *Educational Measurement: Issues and Practice*, 16(4), 14-20. <https://doi.org/10.1111/j.1745-3992.1997.tb00604.x>
- Brennan, R. L. (2000). (Mis)conceptions about generalizability theory. *Educational Measurement: Issues and Practice*, 19(1), 5-10. <https://doi.org/10.1111/j.1745-3992.2000.tb00017.x>
- Brennan, R. L. (2001). *Generalizability theory*. Springer Verlag.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. The Guilford Press.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105. <https://doi.org/10.1037/h0046016>
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63(1), 83-117. <https://doi.org/10.1111/j.1744-6570.2009.01163.x>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Wadsworth.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. John Wiley.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302. <https://doi.org/10.1037/h0040957>
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16(2), 137-163. <https://doi.org/10.1111/j.2044-8317.1963.tb00206.x>
- DeShon, R. P. (2002). Generalizability theory. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations* (pp. 189-220). Jossey-Bass.
- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, 13(3), 230-253. <https://doi.org/10.1037/a0013219>
- Eignor, D. R. (2013). The standards for educational and psychological testing [Professional Ethics & Standards & Liability 3450]. *APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology*, 245-250. <https://doi.org/10.1037/14047-013>
- Finch, W. H., Jr., & French, B. F. (2015). *Latent variable modeling with R*. Routledge/Taylor & Francis Group.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3), 515-533. <https://doi.org/10.1214/06-BA117A>
- Gnambs, T. (2015). Facets of measurement error for scores of the Big Five: Three reliability generalizations. *Personality and Individual Differences*, 84, 84-89. <https://doi.org/10.1016/j.paid.2014.08.019>
- Gorman, C. A., & Rentsch, J. R. (2009). Evaluating frame-of-reference rater training effectiveness using performance schema accuracy. *Journal of Applied Psychology*, 94(5), 1336-1344. <https://doi.org/10.1037/a0016476>
- Guenole, N., Englert, P., & Taylor, P. J. (2003, Jun). Ethnic group differences in cognitive ability test scores within a New Zealand applicant sample. *New Zealand Journal of Psychology*, 32(1), 49-54. <Go to ISI>://000184192100007
- Highhouse, S., Broadfoot, A., Yugo, J. E., & Devendorf, S. A. (2009). Examining corporate reputation judgments with generalizability theory. *Journal of Applied Psychology*, 94(3), 782-789. <https://doi.org/10.1037/a0013934>
- Jackson, D. J. R. (2012). Task-based assessment centers: Theoretical perspectives. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers*. (pp. 173-189). Routledge/Taylor & Francis Group.
- Jackson, D. J. R., & Englert, P. (2011). Task-based assessment centre scores and their relationships with work outcomes. *New Zealand Journal of Psychology*, 40, 37-46.
- Jackson, D. J. R., Kim, S., Lee, C., Choi, Y., & Song, J. (2016). Simulating déjà vu: What happens to game performance when controlling for situational features? *Computers in Human Behavior*, 55, 796-803. <https://doi.org/10.1016/j.chb.2015.10.031>
- Jackson, D. J. R., Michaelides, G., Dewberry, C., Schwencke, B., & Toms, S. (2020). The implications of unconfounding multisource performance ratings. *Journal of Applied Psychology*, 105(3), 312-329. <https://doi.org/10.1037/apl0000434>
- Jackson, D. J. R., Michaelides, M., Dewberry, C., & Kim, Y. (2016). Everything that you have ever been told about assessment center ratings is confounded. *Journal of*

- Applied Psychology*, 101(7), 976-994.  
<https://doi.org/10.1037/ap10000102>
- Jackson, D. J. R., Stillman, J. A., & Atkins, S. G. (2005). Rating tasks versus dimensions in assessment centers: A psychometric comparison. *Human Performance*, 18(3), 213-241. [https://doi.org/10.1207/s15327043hup1803\\_2](https://doi.org/10.1207/s15327043hup1803_2)
- Jackson, D. J. R., Stillman, J. A., & Englert, P. (2010). Task-based assessment centers: Empirical support for a systems model. *International Journal of Selection and Assessment*, 18(2), 141-154.  
<https://doi.org/10.1111/j.1468-2389.2010.00496.x>
- Kleinmann, M., & Köller, O. (1997). Construct validity of assessment centers: Appropriate use of confirmatory factor analysis and suitable construction principles. *Journal of Social Behavior and Personality*, 12(5), 65-84.
- Krause, D. E., & Thornton, G. C., III. (2009, Oct). A cross-cultural look at assessment center practices: Survey results from Western Europe and North America. *Applied Psychology: An International Review*, 58, 557-585.  
<https://doi.org/DOI.10.1111/j.1464-0597.2008.00371.x>
- Kuncel, N. R., & Sackett, P. R. (2014). Resolving the assessment center construct validity problem (as we know it). *Journal of Applied Psychology*, 99(1), 38-47.  
<https://doi.org/10.1037/a0034147>
- Lance, C. E. (2012). Research into task-based assessment centers. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers*. (pp. 218-233). Routledge/Taylor & Francis Group.
- Lance, C. E., Foster, M. R., Nemeth, Y. M., Gentry, W. A., & Drollinger, S. (2007). Extending the nomological network of assessment center construct validity: Prediction of cross-situationally consistent and specific aspects of assessment center performance. *Human Performance*, 20(4), 345-362.  
<https://doi.org/10.1080/08959280701522031>
- Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology*, 89(2), 377-385.  
<https://doi.org/10.1037/0021.9010.89.2.377>
- Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, 7(2), 228-244.  
<https://doi.org/10.1037/1082-989X.7.2.228>
- Lance, C. E., Woehr, D. J., & Meade, A. W. (2007). A Monte Carlo investigation of assessment center construct validity models. *Organizational Research Methods*, 10(3), 430-448. <https://doi.org/10.1177/1094428106289395>
- Le, H., Schmidt, F. L., & Putka, D. J. (2009). The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organizational Research Methods*, 12(1), 165-200.  
<https://doi.org/10.1177/1094428107302900>
- Lievens, F. (2001a). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology*, 86(2), 255-264.  
<https://doi.org/10.1037/0021-9010.86.2.255>
- Lievens, F. (2001b). Assessors and use of assessment centre dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior*, 22(3), 203-221.  
<https://doi.org/10.1002/job.65>
- Liu, X., Rong, J., & Liu, X. (2008). Best linear unbiased prediction for linear combinations in general mixed linear models. *Journal of Multivariate Analysis*, 99(8), 1503-1517.  
<https://doi.org/https://doi.org/10.1016/j.jmva.2008.01.004>
- LoPilato, A. C., Carter, N. T., & Wang, M. (2015). Updating generalizability theory in management research: Bayesian estimation of variance components. *Journal of Management*, 41(2), 692-717.  
<https://doi.org/10.1177/0149206314554215>
- Marcoulides, G. A. (1990). An alternative method for estimating variance components in generalizability theory. *Psychological Reports*, 66(2), 379-386.  
<https://doi.org/10.2466/PRO.66.2.379-386>
- Marcoulides, G. A. (1996). Estimating variance components in generalizability theory: The covariance structure analysis approach. *Structural Equation Modeling*, 3(3), 102-109. <https://doi.org/10.1080/10705519609540045>
- Michalak, R. T., Kiffin-Petersen, S. A., & Ashkanasy, N. M. (2019). I feel mad so I be bad': The role of affect, dissatisfaction and stress in determining responses to interpersonal deviance. *British Journal of Management*, 30(3), 645-667. <https://doi.org/10.1111/1467-8551.12286>
- Murphy, K. R., & DeShon, R. (2000). Interrater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology*, 53(4), 873-900.  
<https://doi.org/10.1111/j.1744-6570.2000.tb02421.x>
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Pulakos, E. D. (1986). The development of training programs to increase accuracy with different rating tasks. *Organizational Behavior & Human Decision Processes*, 38(1), 76-91. [https://doi.org/10.1016/0749-5978\(86\)90027-0](https://doi.org/10.1016/0749-5978(86)90027-0)
- Putka, D. J., & Hoffman, B. J. (2013). Clarifying the contribution of assessee-, dimension-, exercise-, and assessor-related effects to reliable and unreliable variance in assessment center ratings. *Journal of Applied Psychology*, 98(1), 114-133.  
<https://doi.org/10.1037/a0030887>
- Putka, D. J., & Hoffman, B. J. (2014). "The" reliability of job performance ratings equals 0.52. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 247-275). Taylor & Francis.
- Putka, D. J., Lance, C. E., Le, H., & McCloy, R. A. (2011). A cautionary note on modeling multitrait-multirater data arising from ill-structured measurement designs. *Organizational Research Methods*, 14(3), 503-529.  
<https://doi.org/10.1177/1094428110362107>
- Putka, D. J., Le, H., McCloy, R. A., & Diaz, T. (2008). Ill-structured measurement designs in organizational research: implications for estimating interrater reliability. *Journal of Applied Psychology*, 93(5), 959-981.  
<https://doi.org/10.1037/0021-9010.93.5.959>
- Putka, D. J., & Sackett, P. R. (2010). Reliability and validity. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of Employee Selection* (pp. 9-49). Routledge.
- Raykov, T., & Marcoulides, G. A. (2006). Estimation of generalizability coefficients via a structural equation modeling approach to scale reliability evaluation. *International Journal of Testing*, 6(1), 81-95.  
[https://doi.org/10.1207/s15327574ijt0601\\_5](https://doi.org/10.1207/s15327574ijt0601_5)
- Ree, M. J., Carretta, T. R., & Teachout, M. S. (2015). Pervasiveness of dominant general factors in organizational measurement. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 8(3), 409-427.  
<https://doi.org/10.1017/iop.2015.16>



- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Saunders, M. N. K., & Townsend, K. (2016). Reporting and justifying the number of interview participants in organization and workplace research [Research Methods & Experimental Design 2260]. *British Journal of Management*, 27(4), 836-852. <https://doi.org/http://dx.doi.org/10.1111/1467-8551.12182>
- Schmidt, F. L., Viswesvaran, C., & Ones, D. S. (2000). Reliability is not validity and validity is not reliability. *Personnel Psychology*, 53(4), 901-912. <https://doi.org/10.1111/j.1744-6570.2000.tb02422.x>
- Searle, S. R., Casella, G., & McCulloch, C. E. (2006). *Variance components*. Wiley.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Sage.
- Soltani, E., Van Der Meer, R., & Williams, T. M. (2005). A contrast of HRM and TQM approaches to performance management: Some evidence. *British Journal of Management*, 16(3), 211-230. <https://doi.org/10.1111/j.1467-8551.2005.00452.x>
- Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *The American Journal of Psychology*, 18(2), 161-169. <https://doi.org/10.2307/1412408>
- Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5, 1-25. <https://doi.org/10.1146/annurev.clinpsy.032408.153639>
- Taylor, P. J., Keelty, Y., & McDonnell, B. (2002). Evolving personnel selection practices in New Zealand organizations and recruitment firms. *New Zealand Journal of Psychology*, 32, 49-54.
- Tett, R. P., Guterman, H. A., Bleier, A., & Murphy, P. J. (2000). Development and content validation of a "hyperdimensional" taxonomy of managerial competence. *Human Performance*, 13(3), 205-251. [https://doi.org/10.1207/S15327043HUP1303\\_1](https://doi.org/10.1207/S15327043HUP1303_1)
- Thompson, B. (2003). A brief introduction to generalizability theory. In B. Thompson (Ed.), *Score reliability* (pp. 43-58). Sage.
- Thoresen, C. J., & Thoresen, J. D. (2012). How to design and implement a task-based assessment center. In D. J. R. Jackson, C. E. Lance, & B. J. Hoffman (Eds.), *The psychology of assessment centers* (pp. 190-217). Routledge.
- Ward, D. G. (1986). Factor indeterminacy in generalizability theory. *Applied Psychological Measurement*, 10(2), 159-165. <https://doi.org/10.1177/014662168601000206>
- Williams, K. M., & Crafts, J. L. (1997). Inductive job analysis: The job/task inventory method. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement methods in industrial psychology* (pp. 51-88). Davies-Black Publishing.
- Woehr, D. J., Putka, D. J., & Bowler, M. C. (2012). An examination of G-Theory methods for modeling multitrait-multimethod data: Clarifying links to construct validity and confirmatory factor analysis. *Organizational Research Methods*, 15(1), 134-161. <https://doi.org/10.1177/1094428111408616>

#### Corresponding Author

Duncan Jackson

Email: [duncan.jackson@kcl.ac.uk](mailto:duncan.jackson@kcl.ac.uk)