

The Role of Clinical Features in the Diagnostic Reasoning of Psychologists

Cheyenne Gronthos

*This thesis is submitted in partial fulfilment of the Honours degree of Bachelor of  
Psychological Science (Honours)*

School of Psychology

University of Adelaide

September 2021

Word Count: 9,145

**Table of Contents**

<b>List of Figures</b> .....	<b>iv</b>
<b>List of Tables</b> .....	<b>v</b>
<b>Abstract</b> .....	<b>vi</b>
<b>Declaration</b> .....	<b>vii</b>
<b>Contribution Statement</b> .....	<b>viii</b>
<b>Acknowledgements</b> .....	<b>ix</b>
<b>Introduction</b> .....	<b>1</b>
<b>Rationale</b> .....	<b>1</b>
<b>Diagnostic Reasoning in Medicine</b> .....	<b>3</b>
<b>Diagnostic Reasoning in Psychology</b> .....	<b>4</b>
<b>Overview of Expertise</b> .....	<b>7</b>
<i>Generalisable Characteristics</i> .....	<b>7</b>
<i>Feature Utilisation</i> .....	<b>8</b>
<i>The Brunswik Lens Model</i> .....	<b>9</b>
<b>The Present Study</b> .....	<b>11</b>
<b>Method</b> .....	<b>13</b>
<b>Participants</b> .....	<b>13</b>
<b>Study Design</b> .....	<b>14</b>
<b>Materials</b> .....	<b>14</b>
<i>Case Study Development</i> .....	<b>14</b>
<i>Experimental Application</i> .....	<b>18</b>
<i>Data Reduction</i> .....	<b>18</b>
<b>Procedure</b> .....	<b>19</b>

<b>Results .....</b>	<b>21</b>
<b>Overview of Analyses .....</b>	<b>21</b>
<b>Assumption Checking.....</b>	<b>21</b>
<b>Prediction 1 .....</b>	<b>22</b>
<b>Prediction 2.....</b>	<b>23</b>
<b>Prediction 3.....</b>	<b>25</b>
<b>Prediction 4.....</b>	<b>26</b>
<b>Exploratory Findings: Contextual Feature Use .....</b>	<b>28</b>
<b>Discussion.....</b>	<b>31</b>
<b>Overview .....</b>	<b>31</b>
<b>Diagnostic Accuracy and Clinical Experience.....</b>	<b>31</b>
<b>Diagnostic Discriminability Unique and Clinical Experience .....</b>	<b>33</b>
<b>Unique Feature Discriminability and Clinical Experience .....</b>	<b>34</b>
<b>Feature Discriminability and Diagnostic Accuracy .....</b>	<b>35</b>
<b>Contextual Feature Use .....</b>	<b>36</b>
<b>Implications of the Research.....</b>	<b>37</b>
<b>Strengths .....</b>	<b>38</b>
<b>Limitations and Future Directions .....</b>	<b>38</b>
<b>Conclusion .....</b>	<b>40</b>
<b>References .....</b>	<b>41</b>
<b>Appendix A: Study Flyer.....</b>	<b>50</b>
<b>Appendix B: SONA Post.....</b>	<b>51</b>
<b>Appendix C: Information Sheet .....</b>	<b>52</b>
<b>Appendix D: Consent Form .....</b>	<b>54</b>

<b>Appendix E: COVID-19 Screening Questions .....</b>	<b>55</b>
<b>Appendix F: Case Studies .....</b>	<b>56</b>
<b>Appendix G: Regression Analyses R Script.....</b>	<b>59</b>
<b>Appendix H: ANOVA/ANCOVA Analyses R Script .....</b>	<b>63</b>

### List of Figures

Figure 1: The Lens Model diagram.....	10
Figure 2: Example Case Study.....	16
Figure 3: Example Unique Feature.....	16
Figure 4: Example Shared Feature.....	17
Figure 5: Example Contextual Feature.....	18
Figure 6: Interaction Between Years of Clinical Experience and Diagnostic Accuracy.....	23
Figure 7: Interaction Between Diagnostic Discriminability Unique and Years of Clinical Experience.....	24
Figure 8: Scatterplot Demonstrating the Relationship Between Unique Feature Discriminability and Clinical Experience.....	26
Figure 9: Scatterplot Depicting the Relationship Between Unique Feature Discriminability and Diagnostic Accuracy for Each Group.....	27
Figure 10: A Scatterplot Depicting the Relationship Between Diagnostic Accuracy and Diagnostic Discriminability Unique for Novices, Intermediates and Experts.....	28
Figure 11: A Scatterplot with a Fitted Regression Line Displaying the Increase in Contextual Feature Use Associated with Additional Years of Clinical Experience.....	29

**List of Tables**

Table 1: Pearson's Correlation Coefficient for Study Variables.....30

### Abstract

Across health and medical domains, experts rely on idiosyncratic case-based pattern recognition to rapidly and accurately identify significant features that define a case. Understanding how clinical psychologists use features to form a diagnosis can provide valuable insights into changes in diagnostic performance as a function of experience. Previous studies examining the diagnostic accuracy of Clinical Psychologists have demonstrated that practicing Clinical Psychologists are no more accurate at diagnosing mental health conditions than Undergraduate psychology students. This study aimed to explore how the interpretation and use of clinical features develops with experience to facilitate diagnostic reasoning. Undergraduate psychology students ( $n = 24$ ), Clinical Masters students ( $n = 2$ ) and Clinical Psychologists ( $n = 10$ ) were presented with eight mental health case studies. The case studies contained a combination of seven features: those shared between the possible diagnoses and those unique to the primary diagnosis and contextual features. Participants were prompted to give primary and secondary diagnoses for the case studies then asked to rate the extent to which each of the seven features supported the primary and secondary diagnoses. On average, Clinical Psychologists displayed the best diagnostic accuracy. Additionally, tertiary education predicted diagnostic accuracy and the use of unique features but clinical experience was predictive of neither. Rather, clinical experience predicted the use of contextual features (i.e. the character's age or occupation). Future research should extend on these findings using real-life case studies and non-aggregated feature acquisition data.

*Keywords:* features, diagnostic reasoning, Clinical Psychology

**Declaration**

This thesis contains no material which has been accepted for the award of any other degree of diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

Cheyenne Gronthos

September 2021



### **Contribution Statement**

In undertaking this study, I collaborated with my supervisors to refine my research question and methodology, which I had initially developed from my literature review.

Independently, I researched and wrote all sixteen case studies, eight of which were included in the final experiment. My supervisors developed a customised computer program to collect data and assisted me in extracting the data. I was solely responsible for recruiting participants, collecting data and providing remuneration. All sections of this thesis were written independently by myself.

### **Acknowledgements**

I would like to thank my supervisors, Dr Rachel Searston and Dr Daniel Sturman, for all their time and effort dedicated to supporting me this year. From their patience with my endless questions, to their pep talks motivating me to write, their mentorship was instrumental in getting me to the finish line. I enjoyed our weekly meetings and always left feeling energised to persevere with my project.

I would also like to thank Dr Amanda Taylor for contributing ideas for diagnoses to base my case studies on, providing me with invaluable feedback and all her efforts in helping me recruit additional participants. Thank you to Dr Michael Proeve for taking the time to read my case studies and provide valuable feedback on how to further refine the task.

Thank you to all my family and friends for supporting me throughout this year. I want to sincerely thank my Mum and Melanie for listening to me ramble day-to-day and understanding the importance of this year to me. Finally, a thank you to Kahlia for talking to me on the phone at all hours of the day, providing advice and lifting my spirits.

## **The Role of Clinical Features in the Diagnostic Reasoning of Psychologists**

### **Rationale**

Clinical Psychologists are allied health professionals who have undertaken at least eight years of highly specialised training to become experts in mental health. The process of endorsement requires the completion of a six-year sequence: four years of undergraduate study, a minimum two years of post-graduate study (e.g., Masters, Doctorate or combined degree) and a (one to three year) registrar program (APS, 2021). The wide-ranging skillset of Clinical Psychologists includes the assessment and diagnosis of mental health conditions, which are treated using a range of evidence-based therapies and techniques (ACPA, 2021). In accordance with the scientist-practitioner model, psychology is primarily a science-based profession that integrates research and theory with psychotherapy.

Clinical Psychology defines the term *diagnosis* as the process of determining whether an individual is experiencing a mental health condition through careful examination of symptoms and comparison to predefined thresholds, as outlined in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5; APA, 2013; APA, 2015). This task demands the integration of information from multiple sources, including clients' self-reported symptoms, self-reported measures, clinician observations and assessment techniques. Diagnosis also requires psychologists to consider the social, biological and cultural circumstances underlying the presenting problem.

Psycho-diagnosis is a small yet highly significant area of competency in Clinical Psychology. Primarily, it provides the basis of psychotherapy orientation used with a given client, eliminates differential diagnoses and guides treatment decisions (Hill et al., 2017; Love, 2018). For instance, certain mental health conditions require referral to a specialist or specific psychotherapy interventions, such as dialectical behaviour therapy for borderline

personality disorder. Furthermore, diagnosis is a valuable tool for allowing clients to understand and make sense of their experiences.

The Institute of Medicine's *Improving Diagnosis in Health Care* report (2015), highlights that the diagnosis of mental health conditions is a highly dynamic and complex task. Specifically, it outlines issues in psycho-diagnosis, including the over-reliance on self-report measures and clinical observations, which limit the ability to detect cases where a diagnostic error has occurred. This task is further compounded by difficulties in distinguishing between medical and psychological issues, such as those experienced by individuals with chronic pain or unexplained medical symptoms.

Previous studies examining the relationship between clinical experience and diagnostic reasoning in psychotherapy have produced inconsistent findings. While several studies have demonstrated a positive relationship (Brammer, 2002; Wagenaar, 2008), others have demonstrated that Clinical Psychologists are no more accurate than novices (Wittman et al., 2012). For example, one study found no relationship (Wittman & Tollenaar, 2012) and another study found a non-linear U-shaped relationship where Masters students and experienced Clinical Psychologist performed at similar levels (Wittman & van den Bercken, 2007).

The issue of diagnostic performance is increasingly important as subjective accounts from practicing Clinical Psychologists suggest a surge in individuals seeking mental health treatment during the ongoing coronavirus pandemic (COVID-19; Riga, 2020). Psycho-diagnostic performance should have major implications for how Clinical Psychologists are trained in Australia. However, current pathways to accreditation rest on the assumption that experienced psychologists can pass on their expertise through the supervision of provisional psychologists. The underlying belief that trainee psychologists will simply develop diagnostic skills through practice is potentially detrimental to clients as the lack of accurate feedback on

diagnostic performance can result in skill stagnation and/or decline and the perpetuation of errors.

### **Diagnostic Reasoning in Medicine**

While there are inconsistent findings on the diagnostic reasoning of Clinical Psychologists, there is a rich and extensive body of literature concerning medical diagnostic reasoning. Early research in the expert-novice tradition, which compares layman with domain experts, revealed the use of the hypothetico-deductive method by both medical students and clinicians (Barrows et al., 1981; Elstein et al., 1978). That is, within a few minutes of first contact with a patient both groups produced multiple diagnostic hypotheses and collected extra information to confirm or refute their hypotheses. However, clinicians advanced more accurate hypotheses. Medical diagnostic reasoning is characterised by rapid and effective hypothesis generation due to case-based pattern recognition associated with the efficient identification of specific features (both alone or in combination) which define a case (Norman et al., 2018). For example, a study by Gruppen and colleagues (1988) found a rate of 78% diagnostic accuracy in primary care physicians after learning only the patient's main complaint. Critically, the process of pattern recognition is vulnerable to "content specificity", as good diagnostic performance on a previous case is a poor predictor of overall performance, even for similar cases (Elstein et al., 1978; Norman et al., 2018).

Strategy-capturing studies have been employed to analyse how physicians evaluate clinical information when making judgements, yet these studies do not attempt to measure judgement accuracy (Wigton, 1996). Strategy-capturing studies usually consist of case studies with a controlled number of clinical features which are systematically varied in presence and severity. Expert clinicians then rate the individual importance of features present in the case study relative to a given diagnosis. Studies have found that individual and averaged group weightings by clinicians frequently vary from recommendations in the

medical literature (Evans et al., 1995; Kirwan et al., 1984). For example, although level of blood oxygen was emphasised by textbooks as characteristic of pulmonary embolism, internal medicine faculty members ranged in their relative weighting of this feature from 0-90% (Wigton et al., 1986).

Strategy-capturing studies have an advantage in revealing cues that that may not have been otherwise reported during think aloud procedures; either because the clinician is unaware they are attending to it or they perceive the feature as socially unacceptable to report. To illustrate, Rothert et al. (1984) found that a patient's desire for a referral was the most influential factor in general practitioners deciding whether to refer an obese patient to an endocrinologist. Overall, findings in the medical decision-making literature suggest that with experience, psycho-diagnostic reasoning should become increasingly efficient as experts learn to perceive patterns of features. The generalisability of these findings is limited by the fact that objective sources of information (i.e., blood tests and scans) are more readily available for diagnostic decision-making in medicine when compared to psychology (Witteman & Tollenaar, 2012).

### **Diagnostic Reasoning in Psychology**

Expertise in psychotherapy is a hugely contested topic which is limited by conflicting operationalisations of expertise (e.g., clinical experience, reputation, peer nominations and client outcomes) and underlying methodological issues associated with measuring performance in this domain (Hill et al., 2017). For example, both Shanteau (1992) and Tracey et al. (2014) maintain there is little evidence of expertise development in the domain of Clinical Psychology due to a lack of feedback on diagnostic performance and therapy outcomes.

Witteman and van den Bercken (2007) examined the quality of diagnoses developed by psychologists with three levels of experience: Masters students, psychologists with 2-17

years of experience and psychologists with 18+ years of experience. Participants read 10 cases taken from the DSM-4 case book and wrote down their chosen diagnosis.

Unexpectedly, the authors found a non-linear U-shaped relationship where intermediates performed worse than Masters students and clinicians with 18+ years of experience. In addition, there was no significant difference in diagnostic performance between students and psychologists with 18+ years of experience. Similarly, Witteman and Tollenaar (2012) compared the relationship between diagnostic accuracy and free recall of client information across clinicians with various levels of experience using cases from the DSM-4 case book. In the first study, the authors found a non-significant difference in diagnostic accuracy across the groups but in the second study experts were significantly *less* accurate than novices.

In contrast, Wagenaar (2008) compared the diagnostic accuracy of Masters students with or without internship experience and mental health clinicians using a simulated intake interview. Diagnostic accuracy was significantly different between experts and the two groups of Masters students, with experts scoring the highest compared to novices and intermediates. The mixed findings on diagnostic accuracy using different methodologies suggests that psychologists struggle to use clinical features to synthesise a diagnosis but their highly-developed interviewing skills assist them in performing this task.

Brammer (1997) used an artificial intelligence program to examine how level of tertiary education and years of clinical experience effected the diagnostic inquiry of psychologists and graduate psychology students. The program approximated a real-life therapy session by allowing participants to ask questions and simulating client responses using 203 pre-set paragraph answers. Brammer (1997) found that level of tertiary education, years of clinical experience, and number of diagnostic questions significantly predicted greater diagnostic accuracy. In a later study Brammer (2002) re-examined the data using path analysis and identified a direct positive effect between clinical experience and diagnostic

accuracy. Although education had no direct effect, path analysis revealed that advanced tertiary education was associated with an increase in diagnostic questions asked, and a larger number of diagnostic questions asked was associated with increased diagnostic accuracy. This study suggests that experience and tertiary education teach clinicians different skills. Specifically, psychologists were more likely to ask the client contextual questions (e.g., about family, occupation and goals) but this was not associated with their level of university education or diagnostic accuracy.

A meta-analysis of 75 clinical judgement studies by Spengler et al. (2009) found that experience (operationalised as either educational or clinical experience) is positively associated with diagnostic accuracy. Notably, experienced clinicians were better than novices at forming judgements for measures with low-criterion validity ( $d = .22$ ) compared to more naturalistic measures ( $d = .04$ ). The authors posited that the experience-accuracy effect may increase in cases that require a more nuanced understanding of the domain. In a continuation of Spengler et al.'s (2009) meta-analysis, Spengler and Phillips (2015) analysed data from 113 studies and found a small but consistent effect of experience on diagnostic accuracy ( $d = .15$ ), with more experience psychologists demonstrating modest gains in accuracy.

Several authors (Skovholt et al., 1997; Spengler et al., 2009; Wagenaar; 2008) have noted that experience comparisons made in psycho-diagnostic judgement studies frequently capture a limited range of the expert-novice continuum, as most studies compare Masters students to experienced psychologists. Witteman and colleagues (2012) conducted one of the few studies comparing the diagnostic expertise of first-year psychology students, Masters students and Clinical Psychologists. The study used the Cochran–Weiss–Shanteau index, a well-validated empirical measure of expertise, to assess the participants' discriminability and consistency in diagnosing major depression. Participants read 24 vignettes with a controlled



number and type of features (e.g., demographic information, symptoms, and precipitating events). Masters students performed the best and differed significantly from first-year students, while Clinical Psychologists did not differ significantly from first-year students. More importantly, this research focused on diagnostic accuracy in the context of a single diagnosis and did not examine how participants used the clinical information present in the vignettes to make their diagnosis.

## **Overview of Expertise**

### *Generalisable Characteristics*

Expertise has various definitions in the literature due to the influence of two dominant and conflicting approaches, the expert-novice tradition and expert performance approach (Feltovich et al., 2018). In expert-novice studies, expertise is generally associated with increased performance quality due to additional experience and educational attainment (Shanteau, 1992). On the other hand, in the expert performance tradition, experts are the truly elite with “reproducibly superior performance” in a domain (Ericsson et al., 2007, p. 3).

Shanteau (1992) and Kahneman and Klein (2009) concluded the achievement of expertise entails the co-occurrence of two conditions: explicit outcomes in a predictable environment and quality feedback on the accuracy of past actions and decisions. As individuals develop expertise, cognitive differences emerge which mediate task performance, including increased pattern-recognition skills and information-processing capacity (Ericsson, 2014). For example, expert intuition involves the recognition of patterns stored in memory and is characterised by the ability to perceive the structural regularities that define categories and identities within a domain (Kahneman & Klein, 2009). This enables enhanced speed and accuracy in information-processing, and allows experts to initially generate ideal responses to domain-representative tasks (Feltovich et al., 2018; Klein, 1993). In the case of psycho-diagnostic expertise, this increased capacity for pattern recognition would mean that

psychologists should be able to quickly and effectively perceive clusters of symptoms to make a diagnosis.

### *Feature Utilisation*

Several models have been proposed to describe the underlying mechanisms that facilitate pattern-recognition in experts, including the Lens Model (Brunswik, 1955).

*Features* and *cues* are terms used to describe the unconscious automatic relationships between events/objects and features in an environment. They serve as the framework for efficient recognition and response to learned situations (Wiggins et al., 2014a; Wiggins et al., 2014b). The model assumes that successful cue utilisation, regardless of the exact activating features, yields predictable patterns of behaviour in familiar circumstances (Weiss & Shanteau, 2003). Thus, reducing time spent on deliberation during the decision-making process. Feature utilisation serves an adaptive function to assist experts in identifying and attending to critical task features which mediate cognitive processes and performance (Loveday et al., 2014; Shanteau, 1992). Experts have been found to rely on a relatively small number of idiosyncratic features, developed through previous experiences stored in long-term memory, to quickly interpret domain-relevant situations and achieve comparable levels of performance (Brouwers et al., 2017; Lesgold et al., 1988).

Experts with superior feature utilisation can be distinguished from other experts in a domain by an increased capacity to identify and extract task-related features, differentiate between task-relevant features, prioritise information acquisition and discriminate between related and unrelated feature–event relationships (Loveday et al., 2013a; Wiggins, 2014b). Greater feature utilisation is also associated with increased diagnostic performance and superior sustained attention (Loveday et al., 2013a; Loveday et al., 2013b). This has been studied using the EXPERT Intensive Skills Evaluation (EXPERTise 2.0), a validated software shell consisting of five tasks designed to evaluate various aspects of cue utilisation (Wiggins

et al., 2015). The Feature Discrimination Task specifically assesses an individual's ability to discriminate significant environmental information in domain-relevant situations. The task requires individuals to assess a scenario and then rate the utility of each feature in reaching their response. The Feature Association Task requires participants to rate the extent to which a feature is associated with an object or event. Compared to novices, experts display greater variance in ratings for word pair associations through rating a feature as definitely associated with the object or event or rating them as definitely not associated (Watkinson et al., 2018).

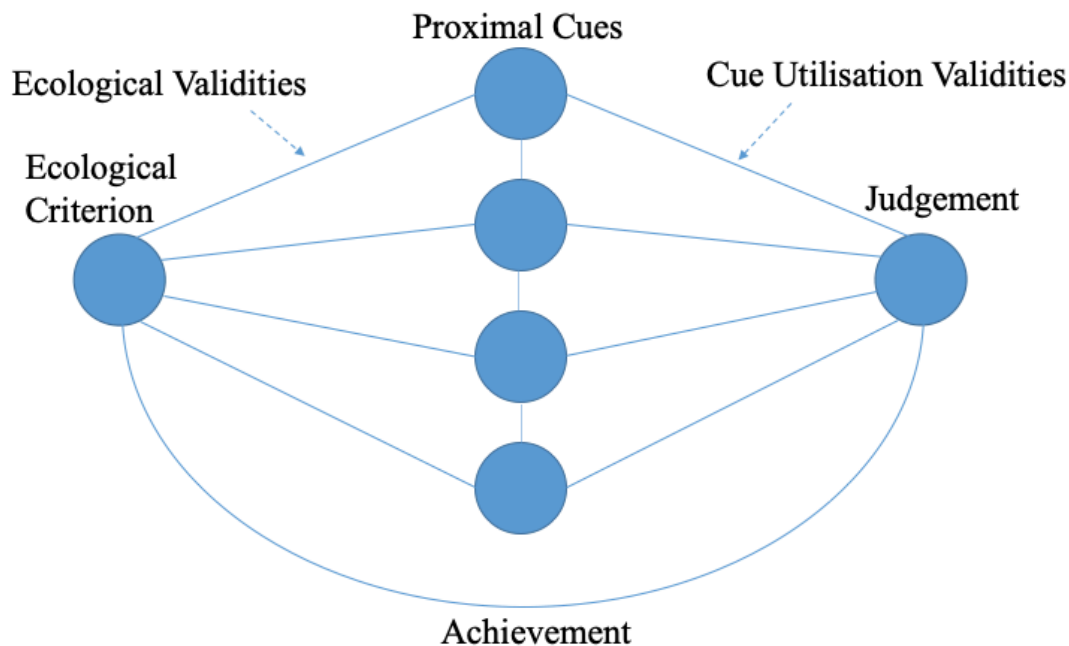
EXPERTise 2.0 has also been applied to feature utilisation in medical diagnostic judgement contexts. For example, Watkinson et al. (2018) found that audiology students displayed significantly increased feature discrimination from the initial to the final stages of their training. Loveday et al. (2013b) found that competent non-experts could be distinguished from experts in paediatric diagnosis using EXPERTise 2.0 as they demonstrated lower feature discrimination than experts.

#### *The Brunswik Lens Model*

The Lens Model (Brunswik, 1955) is a cue-based perception model, which asserts that individuals use probabilistic cues in their environment to make judgements. According to the model, judgements refer to both an aspect of the environment and the relationship between information in the environment and a judgement (Yang & Thompson, 2016). The left side of the diagram below represents ecological validities, the relationship between an ecological criterion (the true state of the situation) and proximal cues in the environment, which are pieces of information that facilitate estimates of the situation. The right side of the model characterises cue utilisation validities, which is the correlation between an individual's judgements and proximal cues in the environment. Achievement represents the relationship between a criterion and an individual's judgement of it. Thus, it reflects the accuracy with which an individual has judged the criterion.

**Figure 1**

*The Lens Model diagram. Adapted from Brunswik (1956).*



*Note.* Ecological criterion is the true state of a situation, while ecological validities are the “correct” cue weights. Proximal cues are information present in the environment and cue utilisation validities are the weighting an individual assigns this information. Judgement is an individual’s assessment of the situation and achievement is the accuracy with which they have judged the true state.

The Lens Model can be applied to understand how Clinical Psychologists make diagnostic decisions. In this context, proximal cues represent symptoms which are disclosed during a session and uncovered through self-report measures, whereas ecological criterion is the client’s “true” diagnosis. Cue validities refer to the strength of relationship between proximal cues and the criterion. Hence, the association between a client’s symptoms and their real diagnosis. The cue utilisation portion of the diagram signifies the relative importance the psychologist assigns different symptoms when making a diagnosis. Judgement refers to the diagnosis the psychologist makes, while ‘achievement’ is the accuracy with which they have judged the criterion (i.e., diagnostic accuracy). For example, to make an accurate diagnosis of

bipolar disorder, a psychologist must apply the most weight to unique symptoms that reflect the true state of the situation (e.g., inflated self-esteem and increased psychomotor agitation) and little weight to symptoms shared with differential diagnoses, such as unipolar depression (e.g., loss of interest in activities and feelings of worthlessness).

### **The Present Study**

This study aimed to explore the relationship between years of clinical experience and the use of clinical features in the diagnosis of mental health conditions. Using an expert-intermediate-novice comparison design, the lab-based task required participants to read eight mental health case studies consisting of seven features. For each case study, participants chose a primary and secondary diagnosis from four options provided. Participants then rated the extent to which features in the case supported their chosen primary and secondary diagnoses. Three types of features were used in the case studies. Some were symptoms outlined in the DSM-5 which are “unique” to the primary diagnosis, while “contextual” features included background information about the character, such as their age or occupation. “Shared” features were symptoms relevant to both the primary and secondary diagnosis.

Based on previous research, the predictions were:

- 1) Clinical Psychologists and Masters students will have better diagnostic accuracy than Undergraduate students. In other words, clinical experience will be a positive predictor of diagnostic accuracy. In accordance with previous meta-analyses (Spengler et al., 2009; Spengler & Pilipis, 2015), it is expected that diagnostic accuracy will increase with clinical experience and tertiary education.
- 2) Clinical Psychologists will show greater variation in their ratings of the unique features between their primary and secondary diagnoses, compared with Masters students

and Undergraduate students. That is, clinical experience will positively predict the extent to which participants discriminate between diagnoses based on their unique features.

Based on findings that experts have an increased capacity to discriminate between feature-object/events (Watkinson et al., 2018; Wiggins, 2014), psychologists are expected to associate unique features with the primary diagnosis and display relatively larger variance in unique features ratings in relation to the primary versus secondary diagnosis.

3) Clinical Psychologists will rate unique features more highly than shared and contextual features in their primary diagnosis, compared with Masters students and Undergraduate students. Thus, clinical experience will positively predict unique feature discriminability. In line with the results of Loveday et al. (2013b) and Watkinson et al. (2018) it is anticipated that expertise as represented by clinical experience will predict increased discrimination of diagnostic features.

4) Unique feature discriminability (prediction 3) and diagnostic discriminability using unique features (prediction 2) will both correspond with higher diagnostic accuracy. As found by previous research (Loveday et al., 2013a; Loveday et al., 2013b), greater feature utilisation—exemplified by the increased discrimination of unique features—will correlate with improved diagnostic accuracy.

## Method

### Participants

A total of 24 Undergraduate Psychology students (18 females), 2 Masters students (1 female), and 10 Clinical Psychologists (9 females) participated in the study. Undergraduate students (“novices”) consisted of first year students enrolled in Psychology 1A and 1B at The University of Adelaide who were recruited through the School of Psychology research participation system (see Appendix B). The two Masters students (“intermediates”) were enrolled in the Masters of Psychology (Clinical) program at The University of Adelaide and were contacted via email and advertisements for the study before classes (Appendix A). Clinical Psychologists (“experts”) were Registered Psychologists with a Clinical Psychology endorsement from Adelaide and across Australia who were recruited using a snowball sampling technique. Participants ranged in age from 17-67 years with median ages of 19, 23.5 and 32 for undergraduate students, comprising Masters students and Clinical Psychologists, respectively. The undergraduate students had completed a median of one semester of tertiary education in psychology whereas Masters students had completed a median of 4.5 years (this equates to a Bachelors and Honours degree in psychology and one semester of the Master’s program) and Clinical Psychologists a median of 6 years (equivalent to Bachelors, Honours and Masters in psychology). Clinical experience amongst the sample of Registered Psychologists ranged from 2 to 20 years with a median of 4 years of experience. Participants were required to be 17 or older and fluent in English. Undergraduate students were granted a full credit for participation while Clinical Psychologists and Masters students went into a draw to win a \$100 Myer/Coles gift voucher purchased by the School of Psychology.

## **Study Design**

This study used a cross-sectional design to examine participants use of clinical features in a mental health diagnostic task as a function of their clinical experience. The task was primarily completed face-to-face in the lab, however, due to the ongoing COVID-19 pandemic and lockdown measures during data collection, 13 participants also completed the task via Zoom. The option to complete the study via Zoom enabled Clinical Psychologists living interstate to participate and reduced the time commitment for local participants. Based on their level of education, participants were assigned to novice, intermediate and expert groups. The primary predictor variables of interest in this study were participants' clinical experience and level of education. The primary performance measures of interest were characteristics of participants' clinical feature ratings, including a measure of their ability to discriminate unique diagnostic features in hypothetical case scenarios and a measure of their ability to use unique features to discriminate the diagnoses in these scenarios. Participants' diagnostic accuracy was also examined in the context of these feature use characteristics.

## **Materials**

### *Case Study Development*

In consultation with a registered Clinical Psychologist, I developed a set of eight case studies *de novo* for the diagnostic task (Appendix F). To generate the case scenarios, I selected several pairs of diagnoses which have similar presentations and are also frequently misdiagnosed. Each case scenario consisted of several clinical features, such as diagnostic criteria and symptoms outlined in the DSM-5, indicative of two plausible diagnoses (e.g., depression and prodromal stage schizophrenia). Three versions of each case scenario were written, which each varied in difficulty based on the number of unique features supporting the primary diagnosis and shared features supporting both the primary and a secondary diagnosis. The number of features in every version was identical. For example, the first



version of the case studies included three features shared across the two most plausible diagnoses and three features unique to the primary diagnosis. The second version included four shared and two unique features, while the third version contained five shared and one unique feature.

In accordance with other psycho-diagnostic reasoning studies (e.g., Witterman & Tollenaar, 2012; Witterman & Van den Bercken, 2007) expert consensus was used as our standard for accurate responses. To ensure high task fidelity, a Clinical Psychologist independently and blindly reviewed the case studies, beginning with the third version of each and worked backwards. Using a thinking aloud procedure, the psychologist selected possible diagnoses based on the features present, provided plausible diagnoses for the forced choice option and gave advice on which version of each case study to use to ensure task sensitivity. A second subject matter expert (also a Clinical Psychologist) blindly reviewed the chosen case studies and suggested further amendments. These amendments included the use of gender pronouns for the characters, enhancing symptom severity for primary diagnoses with lower prevalence rates (i.e., schizophrenia) and increasing the age of characters with stigmatised diagnoses (i.e., conduct disorder). These changes were made to counteract the possibility that Clinical Psychologists may be hesitant to diagnose mental health conditions in the absence of strong evidence due to the stigma associated with certain diagnostic labels.

The final eight selected case studies were revised to produce a version with male and female pronouns for each. Every case study contained different combinations of seven features: those “shared” between the possible diagnoses, those “unique” to the primary diagnosis and “contextual” features (i.e., the character’s age or occupation). The case studies comprised various combinations of the features to minimise ceiling effects in the task. For example, the final version of the Charlie case study (Figure 2, 3, 4 and 5) contains one contextual, two unique and four shared features.

**Figure 2**

*Example Case Study*

Charlie is a second-year university student who lives at home with her family. Ever since starting university last year Charlie has experienced a decrease in energy, which has led her to nap daily after returning home from university. Charlie has also had a significant decline in mood and even hanging out with friends or doing old hobbies does not bring her satisfaction anymore. She now finds her high school friends boring and thinks of them as less intelligent.

Charlie has recently been obsessed with the idea that climate change is a conspiracy created by China to make Australian and US manufacturing non-competitive. She has joined an online forum where she posts frequently about this conspiracy, which has made her nervous that someone from the Chinese government will come looking for her. As the semester has progressed, her grades have suffered due to trouble focusing on readings or lectures. Charlie’s dad thinks she is “just going through a phase”, while her mum is worried about the noticeable decline in her mood and personal hygiene.

Bipolar Disorder	Adjustment Disorder	Prodromal Stage Schizophrenia	Depression
------------------	---------------------	-------------------------------	------------

**What is your primary diagnosis in this case?**

*Note.* Example case study featuring prodromal stage schizophrenia as the primary diagnosis and depression as the secondary. Participants read the case study and selected two diagnoses from the four provided which they believe best account for the symptoms described.

**Figure 3**

*Example Unique Feature*

Charlie is a second-year university student who lives at home with her family. Ever since starting university last year Charlie has experienced a decrease in energy, which has led her to nap daily after returning home from university. Charlie has also had a significant decline in mood and even hanging out with friends or doing old hobbies does not bring her satisfaction anymore. She now finds her high school friends boring and thinks of them as less intelligent.

**Charlie has recently been obsessed with the idea that climate change is a conspiracy created by China to make Australian and US manufacturing non-competitive. She has joined an online forum where she posts frequently about this conspiracy, which has made her nervous that someone from the Chinese government will come looking for her.** As the semester has progressed, her grades have suffered due to trouble focusing on readings or lectures. Charlie’s dad thinks she is “just going through a phase”, while her mum is worried about the noticeable decline in her mood and personal hygiene.

Primary Diagnosis: Prodromal Stage Schizophrenia

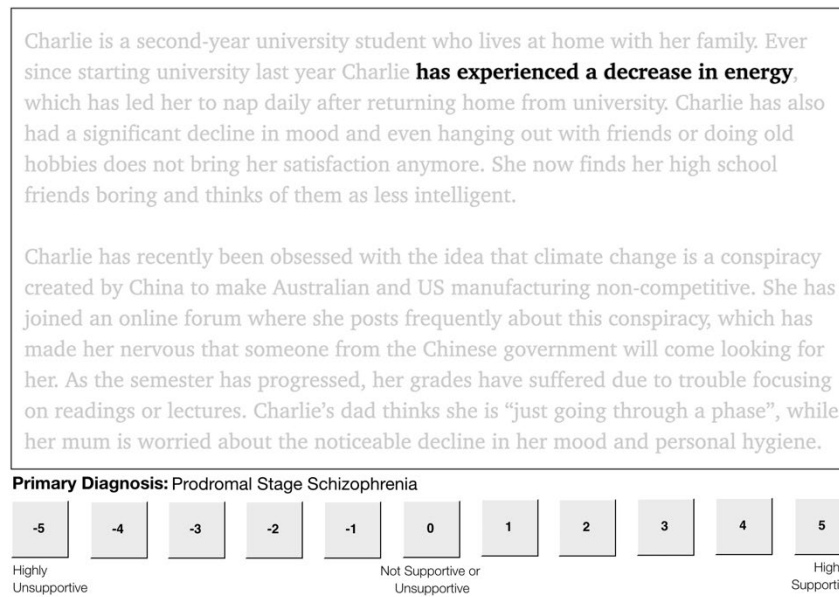
-5	-4	-3	-2	-1	0	1	2	3	4	5
Highly Unsupportive					Not Supportive or Unsupportive					Highly Supportive

**To what extent does the highlighted information support your primary diagnosis?**

*Note.* The highlighted feature is unique to the primary diagnosis of prodromal stage schizophrenia. Participants viewed each feature individually and rated the extent it supported their primary and secondary diagnosis from 5 “highly supportive” to -5 “highly unsupportive”.

**Figure 4**

*Example Shared Feature*



**To what extent does the highlighted information support your primary diagnosis?**

*Note.* The highlighted feature is shared between prodromal stage schizophrenia and depression. Participants rated the importance of this feature in arriving to the primary and secondary diagnoses.

**Figure 5***Example Contextual Feature*

Charlie **is a second-year university student** who lives at home with her family. Ever since starting university last year Charlie has experienced a decrease in energy, which has led her to nap daily after returning home from university. Charlie has also had a significant decline in mood and even hanging out with friends or doing old hobbies does not bring her satisfaction anymore. She now finds her high school friends boring and thinks of them as less intelligent.

Charlie has recently been obsessed with the idea that climate change is a conspiracy created by China to make Australian and US manufacturing non-competitive. She has joined an online forum where she posts frequently about this conspiracy, which has made her nervous that someone from the Chinese government will come looking for her. As the semester has progressed, her grades have suffered due to trouble focusing on readings or lectures. Charlie's dad thinks she is "just going through a phase", while her mum is worried about the noticeable decline in her mood and personal hygiene.

**Primary Diagnosis:** Prodromal Stage Schizophrenia

-5	-4	-3	-2	-1	0	1	2	3	4	5
Highly Unsupportive					Not Supportive or Unsupportive					Highly Supportive

**To what extent does the highlighted information support your primary diagnosis?**

*Note.* Contextual features are not related to either the primary or secondary diagnosis but provide background information on the character.

*Experimental Application*

The diagnostic task was shown to participants using a custom computer program. The application collected and stored participants' demographic information, presented participants with the case scenarios and prompted and recorded their response data. The experiment was presented on 13 and 16-inch monitors with 3072 x 1920 screen resolution.

*Data Reduction*

The individual feature importance data collected during the experiment underwent data reduction. A global feature weighting score was calculated for each participant across the eight case studies based on their average weightings for all three types of features. Mean feature weightings were also calculated for individual participants' unique, shared, and contextual features across the set of case studies. A variance global feature weighting score was then calculated to capture individual variance in combined feature weightings across the

case studies. Variance scores were generated for unique, shared, and contextual feature weightings. Unique feature discriminability scores were generated for each participant to represent the extent to which they discriminated between features unique to the primary diagnosis and other features (i.e., shared and contextual features). Furthermore, diagnostic discriminability scores were calculated for the three types of features to represent the extent that the feature weightings changed from the primary diagnosis to the secondary diagnosis. The diagnostic accuracy data also underwent data reduction. Correct diagnoses were coded as 1 and incorrect diagnoses were coded as 0. Primary diagnostic accuracy was an aggregated score from 0 to 1 which represents an individual's total accuracy in selecting the correct primary diagnosis across the eight case studies. Secondary diagnostic accuracy was also a combined score from 0 to 1 which represents an individual's accuracy in selecting the correct the primary diagnosis in the place of the secondary diagnosis.

### **Procedure**

Ethics approval was granted by the School of Psychology Human Research Ethics Subcommittee at the University of Adelaide (HREC-21/31). Data collection took place from 11<sup>th</sup> of June 2021 to 19<sup>th</sup> of August 2021. Participants were asked COVID-19 screening questions (Appendix E), read an information sheet outlining the basis of the study and signed a consent form (Appendix C and D). In instances where the experiment was conducted via Zoom participants gave verbal consent or via the chat function. A series of demographic questions collected data on the participants' age, gender, level of education and clinical experience. Participants were presented with each individual case study on a laptop screen and were prompted to choose a primary and secondary diagnosis from the four options provided. Only one case (No. 5, Appendix F) included a false diagnosis not contained in the DSM-5 (i.e., psychopathy). Clinical Psychologists were specifically prompted to choose their "best running hypotheses" as none of the case studies contained enough information to make

a diagnosis in accordance with DSM-5 criteria. The selection of a primary and secondary diagnosis was a forced-choice as it was anticipated Clinical Psychologists would refrain from giving a diagnosis unless the character met the full diagnostic criteria. After selecting two diagnoses, participants were then presented each of the seven features in the case study individually and prompted with the question, “To what extent does the highlighted information support your primary diagnosis?”. Features were rated on an 11-point Likert scale from -5 (highly unsupportive) to +5 (highly supportive), with 0 indicating the feature was not supportive or unsupportive. Participants were then prompted to rate the importance of the same feature in relation to the secondary diagnosis before they were presented a new feature. Participants completed the experiment on Zoom via the screen sharing function, which allowed the researcher to share the experiment on the screen and select participants’ chosen responses for them. The gender of the characters, the order of the case studies and the presentation of the features were all semi-randomised to control for the effect of these on diagnosis. Participants were not given feedback on their performance to prevent the answers from being shared.

## Results

### Overview of Analyses

The current study intends to explore differences in clinical feature use and diagnostic accuracy associated with education and clinical experience in psychology. Initially, the planned predictions were to be tested using an ANOVA or ANCOVA (Appendix H). However, we were unable to recruit equal numbers of Masters students ( $n = 2$ ) as compared with practicing Clinical Psychologists ( $n = 10$ ) and Undergraduate students ( $n = 24$ ) and therefore failed to meet the homogeneity of variance assumption of the ANOVA class of tests. Since we had measured clinical experience and level of education using continuous numerical scales, we opted for a regression approach taking full advantage of the variance shown among participants on these measures instead of comparing the three original categorical groups: Undergraduate students, Masters students, Clinical Psychologists. All regression analyses were conducted using a generalised linear model (GLM) assuming a Gaussian distribution (see Appendix G for R script). As we had no pre-specified exclusion criteria, no outliers were excluded from the data set.

### Assumption Checking

Prior to analysis the data were checked for the GLM assumptions, which include normality of residuals, independence of data points, a linear relationship between the response and predictor variables and correct specification of the variance structure. A Shapiro-Wilk test revealed a p-value of .171. Hence, the null hypothesis was not rejected and the residuals were assumed to be normally distributed. Furthermore, due to the nature of the data, independence of data points can be assumed. A scatterplot was used to visually inspect the data and confirm the assumption of a linear relationship between the response and predictor variable. As the assumption of normality was not violated, a Gaussian distribution was selected to represent the variance structure.

**Prediction 1: Clinical experience will be a positive predictor of diagnostic accuracy.**

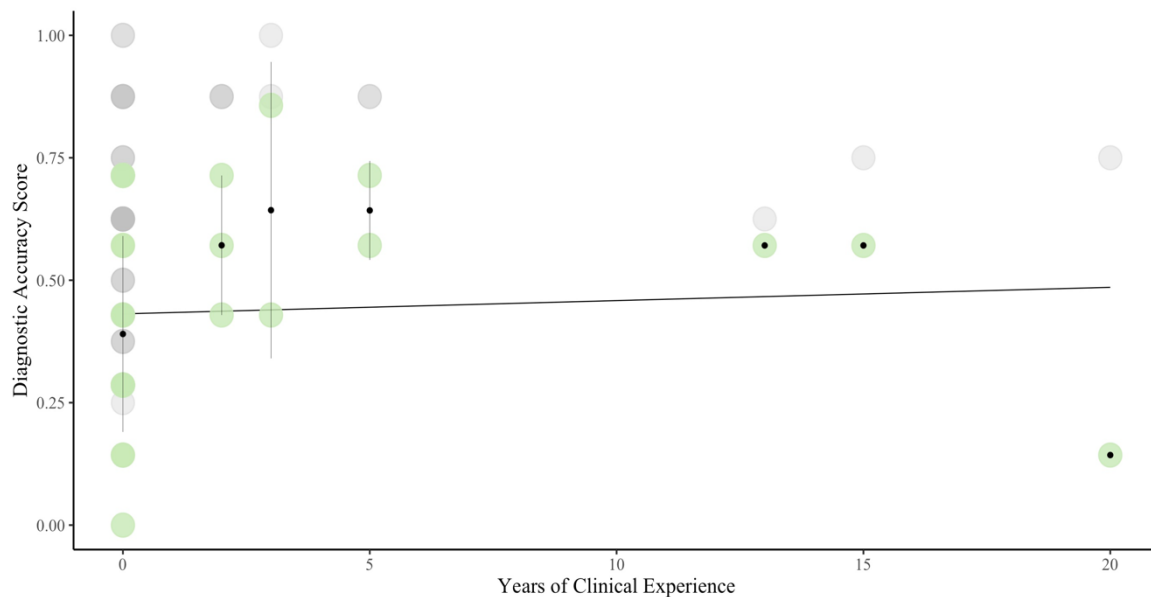
To test the first prediction, a regression analysis was conducted with primary diagnostic accuracy as the outcome variable and level of education and years of clinical experience as the predictor variables. Level of education significantly predicted diagnostic accuracy,  $t(33) = 3.34$ ,  $B = 0.05$ ,  $SE = 0.02$ ,  $p = .002$ , but clinical experience did not,  $t(33) = -1.56$ ,  $B = -0.01$ ,  $SE = 0.01$ ,  $p = .128$ . There was a significant medium positive correlation between primary diagnostic accuracy and level of education,  $r(34) = .45$ ,  $p = .006$ , which indicates that as years of education increase so does diagnostic accuracy. This result suggests that each additional one year of tertiary education is associated with a 5% increase in diagnostic accuracy. In contrast, there is a non-significant very weak positive correlation between primary diagnostic accuracy and experience,  $r(34) = .06$ ,  $p = .732$ . As seen in Figure 6, these results provide support that on average, Masters students ( $M = 0.50$ ,  $SD = 0.30$ ) and Clinical Psychologists ( $M = 0.56$ ,  $SD = 0.20$ ) outperformed first-year psychology students ( $M = 0.38$ ,  $SD = 0.20$ ) in terms of primary diagnostic accuracy. There was also a similar trend for our second measure of accuracy which accounted for participants' primary and secondary diagnostic decisions (if the primary diagnosis was selected as the secondary it was scored as correct using this measure). Experts ( $M = 0.84$ ,  $SD = 0.10$ ) outperformed novices ( $M = 0.65$ ,  $SD = 0.20$ ) and intermediates ( $M = 0.75$ ,  $SD = 0.18$ ). Regression analysis found that education,  $t(33) = 2.94$ ,  $B = 0.04$ ,  $SE = 0.01$ ,  $p = .006$ , was a significant predictor of secondary diagnostic accuracy but experience was not,  $t(33) = -0.90$ ,  $B = -0.01$ ,  $SE = 0.01$ ,  $p = .376$ . Clinical experience had a non-significant correlation with this secondary accuracy measure,  $r(34) = .14$ ,  $p = .409$ , yet university education had a significant moderate positive correlation,  $r(34) = .45$ ,  $p = .006$ . Thus, a one-year increase in tertiary education was associated with a 4% increase in secondary diagnostic accuracy. In all, these results do not support the prediction that clinical experience will predict diagnostic accuracy and instead



suggest years of university education is a better predictor. On average Masters students and Clinical Psychologists outperformed first-year psychology students in diagnostic accuracy as represented by both the primary and secondary accuracy measures.

### Figure 6

#### *Interaction Between Years of Clinical Experience and Diagnostic Accuracy*



*Note.* The scatterplot demonstrates variance in individual primary accuracy scores (in green) and secondary accuracy scores (in grey) with clinical experience. As displayed by the linear regression line, there was no statistically significant increase in diagnostic accuracy with clinical experience.

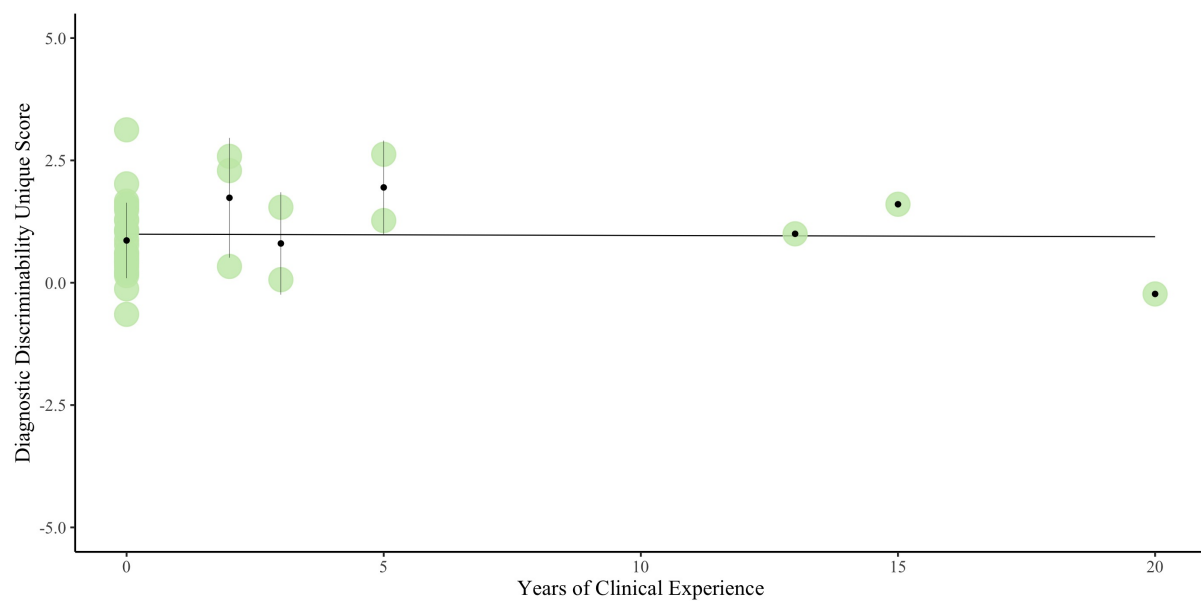
#### **Prediction 2: Clinical experience will positively predict the extent to which participants discriminate between diagnoses based on their unique features.**

A regression analysis was used to examine this prediction with clinical experience and education as predictors of diagnostic discriminability unique, the extent that weightings assigned to unique features change from the primary to the secondary diagnosis. The regression found that clinical experience was not a significant predictor of diagnostic discriminability unique,  $t(33) = -1.51$ ,  $B = -0.05$ ,  $SE = 0.04$ ,  $p = .14$ , yet level of education

was,  $t(33) = 2.54$ ,  $B = 0.17$ ,  $SE = 0.07$ ,  $p = .016$ . According to the regression, a one-year increase in tertiary education is associated with a 17% increase in unique feature weighting variance between the diagnoses. This was evident in the average diagnostic discriminability of unique feature ratings for novices ( $M = 0.81$ ,  $SD = 0.78$ ), intermediates ( $M = 1.47$ ,  $SD = 0.28$ ) and experts ( $M = 1.31$ ,  $SD = 1.02$ ). Furthermore, there was a non-significant moderate positive correlation between diagnostic discriminability unique and education,  $r(34) = .33$ ,  $p = .053$ . In contrast, a non-significant negative correlation was found between clinical experience and diagnostic discriminability based on unique feature ratings,  $r(34) = -.01$ ,  $p = .936$ . Again, these findings fail to support the prediction that clinical experience will predict diagnostic discriminability based on participants' weighting of the unique features in the case scenarios. Rather, tertiary education more accurately predicted the discriminant use of unique features by participants. Figure 7 below depicts the interaction between diagnostic discriminability unique and clinical experience.

### Figure 7

*Interaction Between Diagnostic Discriminability Unique and Years of Clinical Experience*



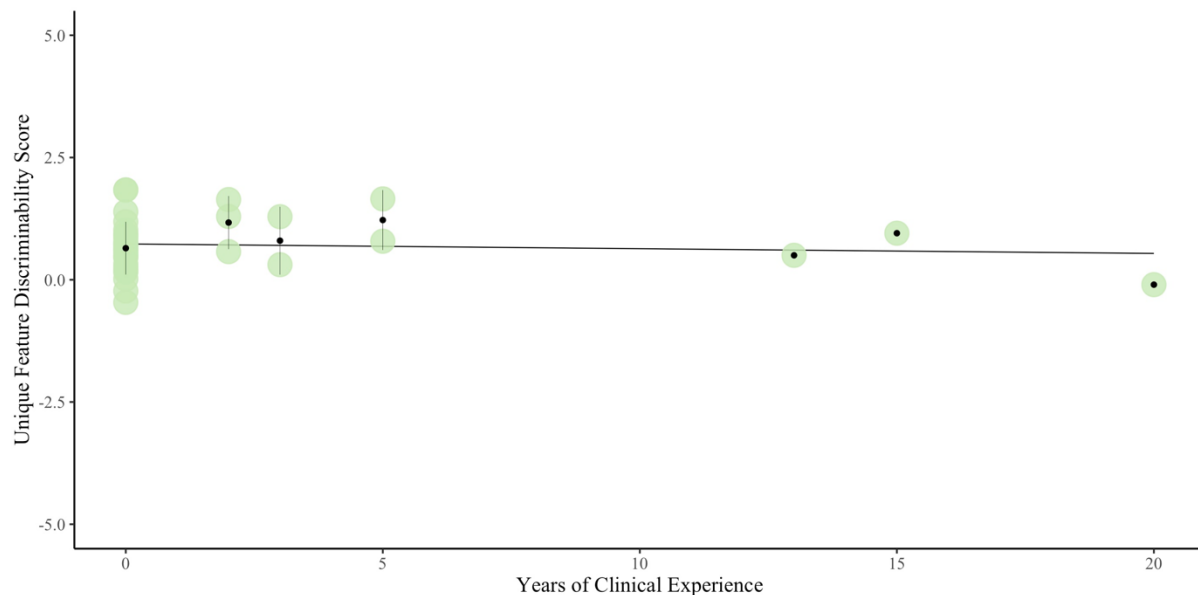
*Note.* Scatterplot displaying differences in diagnostic discriminability unique scores associated with years of clinical experience. The regression line reflects a non-significant association between clinical experience and diagnostic discriminability unique.

**Prediction 3: Clinical experience will positively predict unique feature discriminability.**

Another regression analysis was conducted with level of education and clinical experience as predictors and unique feature discriminability, the extent to which the participants discriminate between features unique to the primary diagnosis and other features (i.e., shared and contextual), as the outcome variable. Years of tertiary education was found to significantly predict unique feature discriminability,  $t(33) = 2.25$ ,  $B = 0.10$ ,  $SE = 0.04$ ,  $p = .031$ , but clinical experience was not,  $t(33) = -1.68$ ,  $B = -0.04$ ,  $SE = 0.02$ ,  $p = .102$ . Thus, when controlling for the effect of clinical experience, each one-year increase in education is associated with a 10% increase in unique feature discriminability. There is a non-significant positive correlation between unique feature discriminability and education,  $r(34) = 0.25$ ,  $p = .134$ , which suggests a suppression effect where education is revealed to be a significant predictor when controlling for the effect of clinical experience. This result would explain the relatively small difference in mean unique feature weightings between novices ( $M = 3.59$ ,  $SD = 0.51$ ), intermediates ( $M = 3.49$ ,  $SD = 0.52$ ) and experts ( $M = 3.62$ ,  $SD = 0.41$ ). Finally, a non-significant correlation was found between unique feature discriminability and clinical experience,  $r(34) = -.08$ ,  $p = .641$ . As with the other predictions, clinical experience does not positively predict unique feature discriminability but years of university education does appear to predict participant discrimination of unique features from shared and contextual features. The interaction between years of clinical experience and unique feature discriminability is captured in Figure 8 below.

**Figure 8**

*Scatterplot Demonstrating the Relationship Between Unique Feature Discriminability and Clinical Experience*



*Note.* The non-significant association between unique feature use and clinical experience is reflected by the regression line.

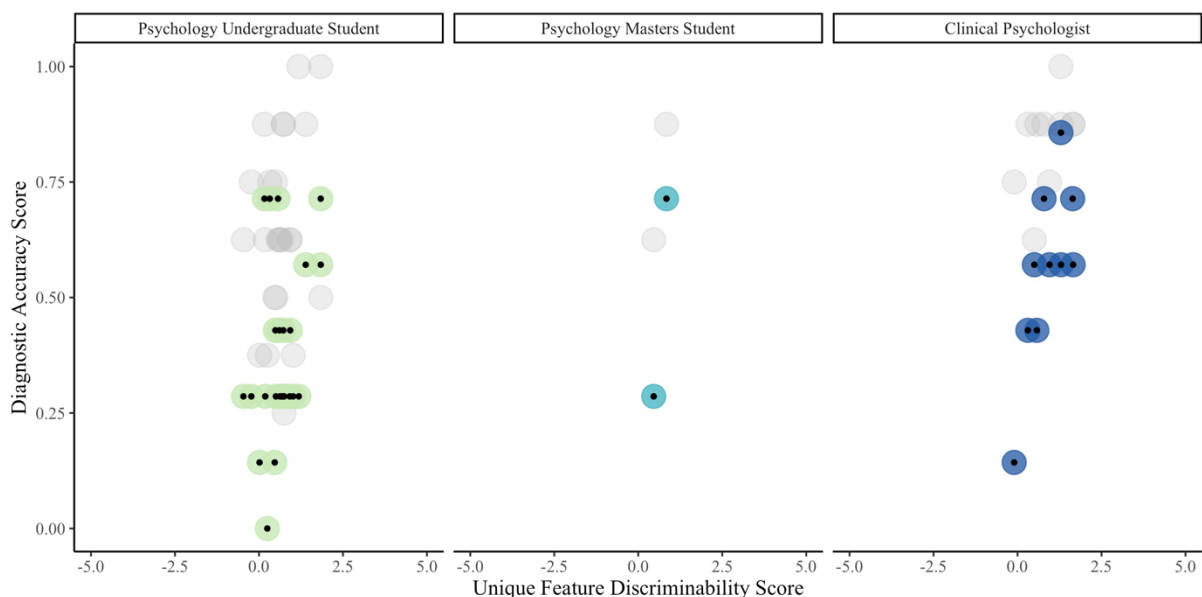
**Prediction 4: Unique feature discriminability (prediction 3) and diagnostic discriminability using unique features (prediction 2) will both correspond with higher diagnostic accuracy.**

A correlation test revealed a moderate-to-strong significant positive correlation between the accuracy of participants primary diagnosis and their discrimination of unique features in the case scenarios,  $r(34) = .51, p = .002$ . As seen in Table 1, diagnostic discriminability unique and primary diagnostic accuracy share a significant and moderate positive correlation,  $r(34) = .43, p = .009$ . This means that as unique feature discriminability and diagnostic discriminability (based on unique feature ratings) increased, so did performance as measured by primary diagnostic accuracy. On average, participants discriminated between the relevance of unique features to the two diagnoses and

discriminated between unique features (compared to shared and contextual features). Improved discriminability in these two facets is associated with increased diagnosis of the correct mental health condition in the first attempt. In contrast, unique feature discriminability,  $r(34) = .33, p = .051$ , and diagnostic discriminability based on unique feature ratings,  $r(34) = .28, p = .095$ , both showed non-significant correlations with secondary accuracy. This result implies a lack of association between diagnostic discriminability unique and accuracy. In this case, as participants weighed unique features greater than other features (contextual and shared), their accuracy increased and participants were more likely to select the correct primary diagnosis for their secondary diagnosis. The significant moderate correlations detailed above support the prediction that unique feature discriminability and diagnostic discriminability based on participants' unique feature ratings correspond with increased primary diagnostic accuracy. Figures 9 and 10 display the correlation between both variables and diagnostic accuracy.

### Figure 9

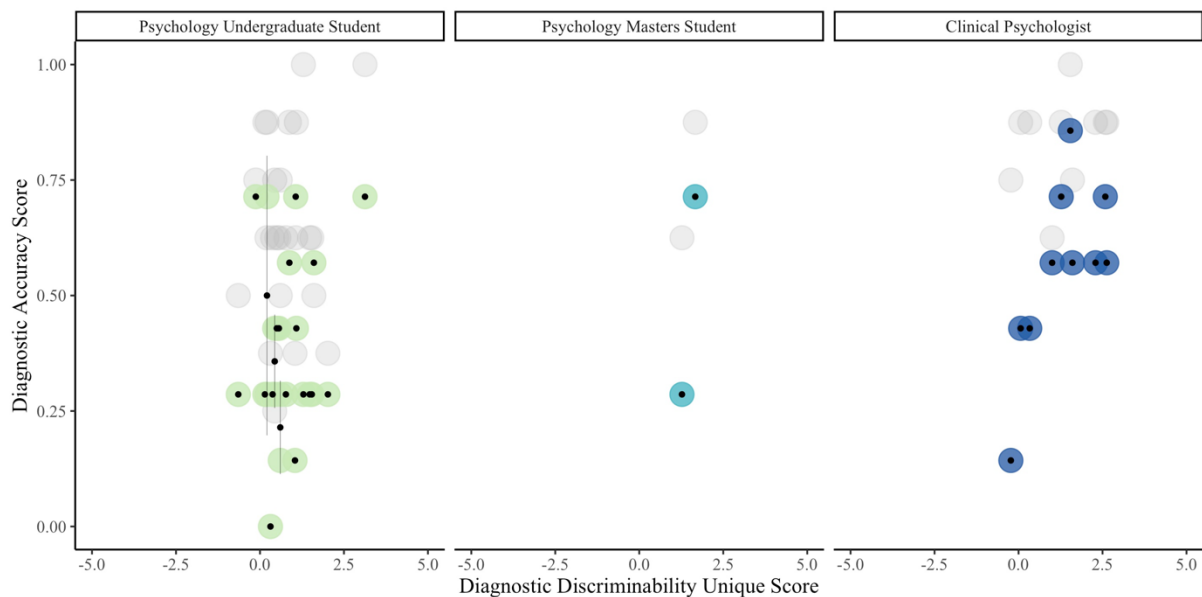
*Scatterplot Depicting the Relationship Between Unique Feature Discriminability and Diagnostic Accuracy for Each Group*



*Note.* The coloured points represent primary accuracy scores while the grey data points reflect secondary accuracy scores. As displayed by the figure, increased unique feature discriminability was associated with improved diagnostic accuracy.

**Figure 10**

*A Scatterplot Depicting the Relationship Between Diagnostic Accuracy and Diagnostic*



*Discriminability Unique for Novices, Intermediates and Experts*

*Note.* Primary accuracy is represented by the coloured data points and secondary accuracy is reflected by the grey data points. The figure reflects that an increased rating of unique features for the primary diagnosis, relative to the unique features ratings for the secondary diagnosis, was associated enhanced diagnostic accuracy.

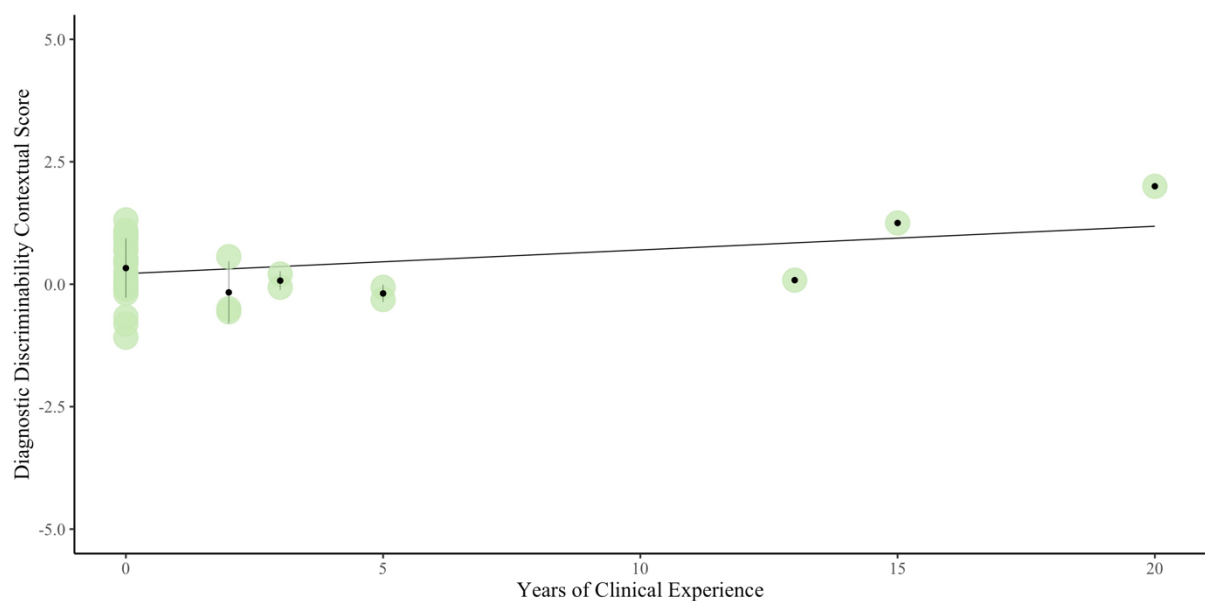
### Exploratory Findings: Contextual Feature Use

Although clinical experience was not predicative of diagnostic accuracy, it was a significant predictor of contextual feature use,  $t(33) = 2.83$ ,  $B = 0.08$ ,  $SE = 0.03$ ,  $p = .008$ . As depicted in Figure 11, a significant moderate positive correlation was uncovered between diagnostic discriminability contextual and experience,  $r(34) = 0.34$ ,  $p = .044$ . This suggests that a one year increase in clinical experience is associated with an 8% increase in contextual

feature use. Novices ( $M = 1.33$ ,  $SD = 1.17$ ) had the highest mean contextual feature weightings followed by intermediates ( $M = 1.75$ ,  $SD = 0.09$ ) and experts ( $M = 0.68$ ,  $SD = 1.16$ ). However, experts ( $M = 0.86$ ,  $SD = 0.77$ ) had the largest average difference between contextual feature weightings relative to the primary and secondary diagnosis compared to novices ( $M = 0.59$ ,  $SD = 0.57$ ) and intermediates ( $M = 0.49$ ,  $SD = 0.56$ ). In contrast, education was not a significant predictor of contextual feature use,  $t(33) = -1.84$ ,  $B = -0.09$ ,  $SE = 0.05$ ,  $p = .074$  and there was a non-significant correlation between the two,  $r(34) = -.05$ ,  $p = .788$ . Finally, a regression analysis revealed that neither clinical experience,  $t(33) = 0.50$ ,  $B = 0.00$ ,  $SE = 0.01$ ,  $p = .619$ , or diagnostic discriminability contextual,  $t(33) = -0.53$ ,  $B = -0.03$ ,  $SE = 0.06$ ,  $p = .597$ , were significant predictors of diagnostic accuracy. As seen in Table 1, primary accuracy has non-significant correlations with both diagnostic discriminability contextual and clinical experience. An unexpected finding was clinical experience is a significant predictive of contextual feature use and will be explored further in the discussion.

### Figure 11

*A Scatterplot with a Fitted Regression Line Displaying the Increase in Contextual Feature Use Associated with Additional Years of Clinical Experience.*



**Table 1***Pearson's Correlation Coefficient for Study Variables*

	1	2	3	4	5	6	7	8	9	10
1. Age	-									
2. Education	.63**	-								
3. Experience	.93**	.57**	-							
4. Accuracy 1	.14	.45**	.06	-						
5. Accuracy 2	.18	.45**	.14	.60**	-					
6. Mean global feature weighting	-.18	-.15	-.18	-.25	-.21	-				
7. Unique feature discriminability	-.12	.25	-.08	.51**	.33	-.10	-			
8. Diagnostic discriminability unique	-.05	.33	-.01	.43**	.28	-.17	.70**	-		
9. Diagnostic discriminability shared	.29	.11	.32	-.21	-.08	-.17	-.11	.26	-	
10. Diagnostic discriminability contextual	.36*	-.05	.34*	-.07	-.13	.11	-.34*	-.08	.32	-

Note. \*  $p < .05$ . \*\*  $p < .01$ .



## **Discussion**

### **Overview**

The current study set out to explore how the use of clinical features changes with experience to facilitate the diagnosis of mental health conditions. Overall, the findings of this study suggest that education was more predictive of diagnostic accuracy than clinical experience. Underlying the enhanced accuracy associated with education was also the increased discriminability of unique features, both from other types of features (i.e., shared and contextual) and between the diagnoses. Unexpectedly, exploratory analyses revealed that clinical experience was associated with contextual feature use, but neither clinical experience nor contextual feature use were associated with diagnostic accuracy. This indicates that with additional years of clinical experience, psychologists increasingly attend to contextual information, such as the character's age or occupation, to make a primary diagnosis.

### **Diagnostic Accuracy and Clinical Experience**

Previous meta-analyses (Spengler et al., 2009; Spengler & Phillips, 2015) have found a small experience-accuracy effect in psycho-diagnosis, particularly on tasks with low criterion validity. Hence, through increased clinical experience, Clinical Psychologists and Masters students were predicted to outperform Undergraduate psychology students in diagnostic accuracy. On average, experts and intermediates had superior diagnostic accuracy. However, this prediction was not supported by the results given that years of clinical experience was not significantly predictive of diagnostic accuracy, as represented by primary or secondary diagnostic accuracy. Rather, like Brammer (1997; 2002), tertiary education was more predictive of both measures of diagnostic accuracy. Brammer (1997; 2002) found that postgraduate education was associated with increased use of diagnostic questions (i.e., about symptoms and alcohol use) but clinical experience was related to asking background questions about the character. The results of the current study appear to reflect the finding

that education and clinical experience teach psychologists to attend to different types of information, with postgraduate education possibly instilling more necessary skills for diagnostic reasoning. Specifically, a one year increase in tertiary education was associated with a 5% increase in primary diagnostic accuracy and a 4% increase in secondary diagnostic accuracy. This means that after six years of psychology education, an individual achieves a 30% increase in diagnostic accuracy. This would have important implications for treatment planning and patient outcomes, particularly in contexts where clinicians are required to provide a diagnosis (e.g., for insurance and legal reasons).

One factor that may have influenced the results and made education more predictive than experience is the fact that most of the Clinical Psychologists who participated in the study were only recently endorsed (within the last 2 to 4 years). In the domain of medicine, expertise development is characterised by the knowledge restructuring process of knowledge encapsulation, a prolonged process involving the abbreviation of reasoning lines and the integration of lower-order concepts into higher-order concepts (Boshuizen et al., 2020). During the process of knowledge encapsulation, an inverted-U-shaped curve or “intermediate effect” emerges which reflects an increase in recall performance for bio-medical information with up to six years of training (Schmidt & Boshuizen, 1992). However, recall performance decreases with further experience as experts primarily use macro-concepts and rarely refer to underlying pathophysiological mechanisms. Intermediates have been found to use more technical knowledge whereas experts largely rely on case-based knowledge, which is supplemented with technical knowledge when facing novel situations. Consistent with the process of knowledge encapsulation, psychotherapy experts have been found to possess complex and abstract knowledge bases, consisting of more domain-specific concepts (Boshuizen et al., 2020; Marsh & Ahn, 2012). In classifying cases of major depression, Witteman et al. (2012) found an intermediate effect in participants’ consistency and

discriminatory precision of DSM categories. Thus, suggesting that with higher levels of expertise in Clinical Psychology, diagnostic categories tend to blur. The tendency for experts to gloss over features of a problem (both surface characteristics and details) due to knowledge automatization and pattern-recognition has also been demonstrated in other domains (Chi, 2006). It is possible that experienced Clinical Psychologists may have more trouble distinguishing diagnoses from similar diagnostic categories, which was indeed the case for some pairs of diagnoses used in the case studies, like acute stress disorder and post-traumatic stress disorder. Furthermore, the greater influence of tertiary education on diagnostic accuracy may have reflected an intermediate effect in the results which would have been revealed with an ANOVA class analysis. In Australia, due to an emphasis on the science-practitioner approach in undergraduate psychology degrees, most psychologists only get an in-depth education on DSM-5 diagnostic criteria during their Master's degree (APS, 2021). Therefore, recently endorsed Clinical Psychologists may possess more technical knowledge of DSM-5 categories as their domain-relevant knowledge structures have not yet been fully encapsulated.

### **Diagnostic Discriminability Unique and Clinical Experience**

Through measuring variance as a proportion of response latency, previous research has found that experts have an increased capacity to discriminate between feature-object/events (Carrigan et al., 2020; Watkinson et al., 2018; Wiggins, 2014). In line with this research, Clinical Psychologists were expected to weight unique features as more important to the primary diagnosis than the secondary diagnosis and display a relatively larger variance in unique features ratings between the diagnoses. In other words, clinical experience was expected to positively predict the extent to which participants discriminate between diagnoses based on their unique features. This predication was not supported since clinical experience was not significantly related to diagnostic discriminability unique. However, a one year

increase in tertiary education was associated with a 17% increase in unique feature weighting variance between the diagnoses. Hence, with more university education, participants increasingly displayed stronger association between unique features and the primary diagnosis and weaker associations between unique features and other diagnoses. In the context of previous feature utilisation research, concepts that co-occur in working memory become strongly associated and form feature–event relationships (Wiggins, 2006; Wiggins & Bollwerk, 2006; Wiggins & O’Hare, 2003). This association has an increased likelihood of being stored in long-term memory and results in a greater capacity to retrieve related concepts, depending of the strength of the relationship and the specificity of the association between the concepts (Morrison et al., 2013). In Masters programs, the provision of readily available and quality feedback on diagnostic performance allows students to form stronger feature-object/event associations between symptoms and mental health conditions. In contrast, clinical practice in psychology has been specifically highlighted as a “wicked” learning environment where feedback is delayed, irregular or non-existent which can cause skill stagnation and decline (Shanteau, 1992; Tracey et al., 2014).

### **Unique Feature Discriminability and Clinical Experience**

According to the lens model (Brunswik, 1955), behaviour is guided by the judgements an individual makes based on features in the environment and the meaning assigned to them. This model assumes that perceiving familiar features in a scenario yields predictable patterns of behaviour. Based on previous performance on the EXPERTise 2.0 Feature Discrimination Task in medical domains (Loveday et al., 2013b & Watkinson et al., 2018), expertise as exemplified by clinical experience was expected to predict a higher capacity for feature utilisation through the increased discrimination of diagnostic features. Experts were anticipated to have relatively greater unique feature discriminability, which is the capacity to discriminate features unique to the primary diagnosis from shared and contextual features.

This prediction was not supported by the findings given that clinical experience did not predict unique feature discriminability. Yet, with more education, participants rated features unique to the primary diagnosis as higher than shared and contextual features. Specifically, each one year increase in tertiary education was associated with a 10% increase in unique feature discriminability. This aligns with Brammer's (1997; 2002) finding that postgraduate education was associated with an increased use of diagnostic questions (i.e. about symptoms and alcohol use). In accordance with the Lens Model, participants in both studies perceived familiar features (i.e., symptoms) which defined the cases and discriminated them from less significant features. Thus, demonstrating expertise through the capacity to identify the structural regularities in a domain and discriminate relevant features (Kahneman & Klein, 2009; Weiss & Shanteau, 2003).

### **Feature Discriminability and Diagnostic Accuracy**

Previous literature has suggested that increased feature use, involving the enhanced discrimination of unique features, is associated with consistent and accurate diagnostic evaluations (Loveday et al., 2013a; Loveday et al., 2013b; Wiggins et al., 2014a). Therefore, the discrimination of unique features and differentiation of the primary diagnosis based on unique features were predicted to correspond with diagnostic accuracy. The findings supported the prediction that, as unique feature discriminability and diagnostic discriminability increase, so does performance on primary diagnostic accuracy. This was anticipated in accordance with previous literature, specifically, a study of family physicians found that number of "critical cues" requested (i.e., diagnostic cues like unique features) was a significant predictor of diagnostic accuracy (Kostopoulou et al. 2008). Depending on the case study scenario, each additional critical cue increases the odds of arriving at the correct diagnosis by 1.3 to 7.5 times. Furthermore, greater weighting of diagnostic features and a relatively lower rating of irrelevant features is highly predictive of decision performance in a

variety of domains (Wiggins & O'Hare, 2003). Neither types of discriminability were significantly correlated with secondary diagnostic accuracy because they predicted increased diagnostic performance in the first attempt. Diagnostic discriminability unique does not correlate with improved secondary diagnostic accuracy because it captures variance in the unique feature rating between the primary and secondary diagnosis.

### **Contextual Feature Use**

Exploratory analysis of the data revealed that clinical experience but not tertiary education was predictive of the discrimination of contextual features from other features (i.e., shared and unique features). Thus, in choosing the primary diagnosis, a one year increase in clinical experience was associated with an 8% increase in contextual feature use. Notably, neither clinical experience nor diagnostic discriminability were significant predictors of diagnostic accuracy. While unanticipated, this finding is logical when considering previous research. In a simulated therapy session with an artificial intelligence program, Brammer (1997) found that clinical experience significantly predicted the use of questions to gather background information, specifically related to the client's psycho-social history, family, goals and occupation. This study found a direct experience-accuracy effect, yet the propensity to gather non-diagnostic information was not associated with diagnostic accuracy.

In addition, it is possible that the increased use of contextual features associated with clinical experience is a form of "vicarious functioning" where multiple partial cues with limited validity and reliability are used to make a judgement in uncertain environments (Brunswik, 1956). Dhimi and Mumpower (2018) posit that vicarious functioning is vital in medical decision-making contexts where clients can present with dynamic and changing symptoms which differ from those presented by other clients with the same condition, much like the case in Clinical Psychology.

A subject matter expert (i.e., Clinical Psychologist) suggested that contextual feature use may be associated with clinical experience because it is a critical aspect of case formulation. Case formulation is an integrated running hypothesis of the causes, precipitants, and maintaining forces underlying a client's presenting problem (Eells et al., 1998). The integration of this information is also critical for developing treatment plans, which are positively related to level of expertise as expert treatment plans are closely related to case formulation, diagnosis, and treatment options (Caspar et al., 2004; Eells & Lombart, 2010). These findings highlight that the duties of Clinical Psychologists are multifaceted and diagnosis is by no means the only task they perform.

### **Implications of the Research**

The findings suggest that performance in diagnostic reasoning steadily increases with further tertiary education but not with clinical experience. As proposed by Boshuizen et al. (2020), these findings suggest a misalignment in psychology between knowledge learned in tertiary education institutions and task requirements that arise in the workplace. The current study is contributing to the growing literature on the diagnostic performance of Clinical Psychologists. This study is unique because it is the first to specifically examine how psychologists use clinical features to make diagnoses, rather than just assessing the accuracy of diagnostic judgements and case formulations (Boshuizen et al., 2020). In line with medical strategy-capturing studies, this study found that clinicians attend to non-diagnostic information when making diagnostic decisions. The findings of the present study also indicate that psychologists perceive patterns of features to make a diagnosis, through the association of distinctive symptoms with a certain diagnosis and the discrimination of less relevant symptoms. Thus, providing support that psychologists display recognition of patterns stored in memory, a facet of expertise which psychologists have previously been hypothesised to lack (Shanteu, 1992; Tracey et al., 2014).

Understanding psycho-diagnostic reasoning and feature use is critical in training future Clinical Psychologists and designing ongoing professional development programs to prevent skill stagnation or decline in practicing Clinical Psychologists. This is important because previous research has found that increased clinical experience heightens confidence in diagnostic decisions (Tracey et al., 2014). However, this does not correspond with an increase in accuracy and may prevent Clinical Psychologists from acting to improve their performance.

### **Strengths**

There are several strengths of this study that are associated with methodological decisions made throughout the research process. For example, the validation of case studies by two independent Clinical Psychologists and the use of expert consensus to determine accurate responses resulted in high-quality stimuli within the constraints of an Honours project. The repeated measures design of the study also increased the statistical power available to detect an effect, especially given the small sample size. Finally, the decision to analyse the data using regression analyses capitalised on the variation in participants' clinical experience. Instead of collapsing variance like most studies in this domain by using ANOVA or ANCOVA tests (Wittelman & van den Bercken 2007; Wittelman & Tollenaar 2012), variance was explored from zero to 20 years of clinical experience.

### **Limitations and Future Directions**

The current study has several limitations which constrain the generalisability of the results, but offer avenues for future research. Primarily, it was difficult to a sample from specialist populations, which resulted in this study having a small sample size. While Clinical Psychologists were readily recruited via snowball sampling, Masters students were unexpectedly challenging to recruit despite multiple efforts to find participants at the University of Adelaide and other universities. Due to the small sample of Masters students,



the data were analysed using regression analyses rather than the originally planned analyses. The small sample would also have lowered the statistical power available to detect an effect, especially given the literature indicates a small experience-accuracy effect (Spengler et al., 2009; Spengler & Phillips, 2015). Future research, unconstrained by the time limitations imposed during Honours, could recruit a larger and more diverse sample of participants. This would increase statistical power to detect the true population effect and provide a better understanding of how both feature use and diagnostic accuracy change with tertiary education.

Although the use of case studies provided experimental control whereby participants were exposed to a controlled number of features, it is artificial to the task of diagnosis. The use of a low-fidelity task posed the issue that there was no ground truth to the diagnoses, especially since the case studies were not written by a subject matter expert, but a student researcher. While each case study included symptoms outlined in the DSM-5 and real-life case studies published online, there was not enough information in any to make a full diagnosis in accordance with DSM-5 criteria. The use of case studies did not account for the common occurrence of co-morbid mental health conditions. Future research could address these methodological issues using real-life case studies or simulated interviews to explore diagnostic decision-making in a more naturalistic setting. These changes would fill the gap in the literature of how Clinical Psychologists use features in real-world contexts to make diagnoses and would have more applications to psychology training programs.

Finally, a major limitation of the present study is the use of aggregating feature data to assess differences in feature use associated with experience and education. Aggregation of the research data enabled inferences to be made about how participants used different types of features. However, previous research has revealed that across experts there are individual differences in the use and predictive validity of individual features (Wiggins, 2014; Wigton,

1996). Through aggregating feature data, the qualitative differences in the feature-weighting profiles of experts can result in misrepresentative data. For example, examining aggregated unique feature weightings across participants could be misleading because participants may rate unique features as highly important in arriving at the wrong primary diagnosis. Instead of directly examining the individual features used by participants, several researchers have circumvented this issue by exploring the behavioural patterns participants exhibit in acquiring information (Loveday et al., 2013a, Morrison et al., 2013). Future studies could explore how Clinical Psychologists use features in a more holistic sense by examining information acquisition in this population and analysing non-aggregated data.

### **Conclusion**

The aim of the present study was to explore the influence of clinical experience on feature use in diagnostic decision-making. Diagnostic accuracy, both primary and secondary, increased with education but not clinical experience. Education also predicted the discrimination of unique features, from other types of features and between diagnoses, both of which were positively correlated with diagnostic accuracy. Notably, clinical experience only predicted contextual feature, yet neither were associated with diagnostic accuracy. Overall, this research contributes to the growing psycho-diagnostic reasoning and psychology expertise literature. As Clinical Psychologists become increasingly sought out during the ongoing pandemic, future studies should further explore diagnostic reasoning using real-life case studies and non-aggregated feature acquisition data.

### References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/https://doi.org/10.1176/appi.books.9780890425596>
- American Psychological Association. (2015). *APA Dictionary of Psychology* (2 ed.). American Psychological Association. <https://doi.org/10.2307/j.ctv1chrw2d>
- Australian Clinical Psychology Association. (2021). *What is a Clinical Psychologist?* <https://acpa.org.au/what-is-a-clinical-psychologist/>
- Australian Psychological Society. (2021). *Study pathways*. <https://psychology.org.au/training-and-careers/careers-and-studying-psychology/studying-psychology/study-pathways>
- Barrows, H. S., Norman, G. R., Neufeld, V. R., & Feightner, J. W. (1981). The clinical reasoning of randomly selected physicians in general medical practice. *Clinical and Investigative Medicine*, 5(1), 49-55.
- Boshuizen, H. P. A., Gruber, H., & Strasser, J. (2020). Knowledge restructuring through case processing: The key to generalise expertise development theory across domains? *Educational Research Review*, 29(100310), 1-31. <https://doi.org/https://doi.org/10.1016/j.edurev.2020.100310>
- Brammer, R. (1997). Case Conceptualization Strategies: The Relationship Between Psychologists' Experience Levels, Academic Training, and Mode of Clinical Inquiry. *Educational Psychology Review*, 9(4), 333-351. <https://doi.org/https://doi.org/10.1023/A:1024794522386>
- Brammer, R. (2002). Effects of Experience and Training on Diagnostic Accuracy. *Psychological Assessment*, 14(1), 110-113. <https://doi.org/https://doi.org/10.1037/1040-3590.14.1.110>

- Brouwers, S., Wiggins, M. W., Griffin, B., Helton, W. S., & O'Hare, D. (2017). The role of cue utilisation in reducing the workload in a train control task. *Ergonomics*, *60*(11), 1500-1515. <https://doi.org/10.1080/00140139.2017.1330494>
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193-217. <https://doi.org/10.1037/h0047470>
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed. ed.). University of California Press.
- Carrigan, A., Magnussen, J., Georgiou, A., Curby, K., Palmeri, T. J., & Wiggins, M. W. (2020). Differentiating Experience From Cue Utilization in Radiological Assessments. *Human Factors*, *63*(4). <https://doi.org/10.1177/0018720820902576>
- Caspar, F., Berger, T., & Hautle, I. (2004). The right view of your patient: A computer assisted, individualized module for psychotherapy training. *Psychotherapy: Theory, Research, Practice, Training*, *41*(2), 125-135. <https://doi.org/10.1037/0033-3204.41.2.125>
- Chi, M. T. H. (2006). Two Approaches to the Study of Experts' Characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (1st ed., pp. 21-30). Cambridge University Press. <https://doi.org/10.1017/CBO9780511816796.002>
- Committee on Diagnostic Error in Health, B. o. H. C. S., Institute of Medicine. (2015). *Improving Diagnosis in Health Care* (J. R. Ball, B. T. Miller, & E. P. Balogh, Eds.). The National Academies Press. <https://doi.org/10.17226/21794>
- Dhami, M. K., & Mumpower, J. L. (2018). Kenneth R. Hammond's contributions to the study of judgment and decision making. *Judgment and Decision Making*, *13*(1), 1-22.

- Eells, T. D., Kendjelic, E. M., & Lucas, C. P. (1998). What's in a case formulation? Development and use of a content coding manual. *The Journal of Psychotherapy Practice and Research*, 7(2), 144-153.
- Eells, T. D., & Lombart, K. G. (2003). Case Formulation and Treatment Concepts Among Novice, Experienced, and Expert Cognitive-Behavioral and Psychodynamic Therapists. *Psychotherapy research*, 13(2), 187-204. <https://doi.org/10.1093/ptr/kpg018>
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Harvard University Press.
- Ericsson, K. A. (2014). Why expert performance is special and cannot be extrapolated from studies of performance in the general population: A response to criticisms. *Intelligence*, 45, 81-103. <https://doi.org/10.1016/j.intell.2013.12.001>
- Ericsson, K. A., Roring, R. W., & Nandagopal, K. (2007). Giftedness and evidence for reproducibly superior performance: an account based on the expert performance framework. *High Ability Studies*, 18(1), 3-56. <https://doi.org/10.1080/13598130701350593>
- Evans, J. S., Harries, C., Dennis, I., & Dean, J. (1995). General practitioners' tacit and stated policies in the prescription of lipid lowering agents. *British Journal of General Practice*, 45(390), 15-18.
- Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2018). Studies of Expertise from Psychological Perspectives: Historical Foundations and Recurrent Themes. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316480748.006>
- Gruppen, L. D., Woolliscroft, J. O., & Wolf, F. M. (1988). The contribution of different components of the clinical encounter in generating and eliminating diagnostic

hypotheses. In D. S. Dabney (ed.), *Research in medical education: Proceedings of the annual conference on research in medical education* (pp. 242-247). Washington, DC: Association of American Medical Colleges.

Hill, C. E., Spiegel, S. B., Hoffman, M. A., Kivlighan, D. M., & Gelso, C. J. (2017).

Therapist Expertise in Psychotherapy Revisited. *The Counseling Psychologist*, 45(1), 7-53. <https://doi.org/10.1177/0011000016641192>

Kahneman, D., & Klein, G. (2009). Conditions for Intuitive Expertise: A Failure to Disagree. *American Psychologist*, 64(6), 515-526.

<https://doi.org/https://doi.org/10.1037/a0016755>

Kirwan, J. R., Chaput de Saintonge, D. M., Joyce, C. R., & Currey, H. L. (1984). Clinical judgment in rheumatoid arthritis. III. British rheumatologists' judgments of 'change in response to therapy'. *Annals of the Rheumatic Diseases*, 43(5), 686-694.

<https://doi.org/10.1136/ard.43.5.686>

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 138-147). Ablex Publishing.

Kostopoulou, O., Oudhoff, J., Nath, R., Delaney, B. C., Munro, C. W., Harries, C., & Holder, R. (2008). Predictors of Diagnostic Accuracy and Safe Management in Difficult Diagnostic Problems in Family Medicine. *Medical Decision Making*, 28(5), 668-680.

<https://doi.org/10.1177/0272989X08319958>

Lesgold, A., Rubinson, H., Feltovich, P. J., Glaser, R., Klopfer, D., & Wang, Y. (1988).

Expertise in a Complex Skill: Diagnosing X-ray Pictures. In M. T. H. Chi, R. Glaser, & M. Farr (Eds.), *The Nature of Expertise* (pp. 311-342). Earlbaum.

Love, A. (2018, February 1). The diagnostic dilemma. *InPsych*, 40(1).

<https://psychology.org.au/for-members/publications/inpsych/2018/feb/the-diagnostic-dilemma>

Loveday, T., Wiggins, M. W., Harris, J. M., O'Hare, D., & Smith, N. (2013a). An Objective Approach to Identifying Diagnostic Expertise Among Power System Controllers.

*Human Factors* 55(1), 90-107. <https://doi.org/10.1177/0018720812450911>

Loveday, T., Wiggins, M. W., & Searle, B. J. (2014). Cue Utilization and Broad Indicators of Workplace Expertise. *Journal of Cognitive Engineering and Decision Making*, 8(1), 98-

113. <https://doi.org/10.1177/1555343413497019>

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2013b). The Capability of Static and Dynamic Features to Distinguish Competent From Genuinely Expert

Practitioners in Pediatric Diagnosis. *Human Factors*, 55(1), 125-137.

<https://doi.org/10.1177/0018720812448475>

Marsh, J. K., & Ahn, W.K. (2012). Memory for Patient Information as a Function of

Experience in Mental Health: Memory for patient information. *Applied Cognitive*

*Psychology*, 26(3), 462-474. <https://doi.org/10.1002/acp.2832>

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2013). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues.

*Journal of Cognitive Engineering and Decision Making*, 7(2), 211-226.

<https://doi.org/https://doi.org/10.1177/1555343412459192>

Norman, G. R., Grierson, L. E. M., Sherbino, J., Hamstra, S. J., Schmidt, H. G., & Mamede,

S. (2018). Expertise in Medicine and Surgery. In K. A. Ericsson, R. R. Hoffman, A.

Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert*

*Performance* (2nd ed., pp. 331-355): Cambridge University Press.

- Riga, R. (2020, October 4). *Queensland doctors and psychologists facing a 'tsunami of mental health presentations' after months of coronavirus limitations*. ABC News.  
<https://www.abc.net.au/news/2020-10-04/covid-19-qld-mental-health-long-wait-times-doctors-struggling/12726854>
- Rothert, M. L., Rovner, D. R., Elstein, A. S., Holzman, G. B., Holmes, M. M., & Ravitch, M. M. (1984). Differences in Medical Referral Decisions for Obesity among Family Practitioners, General Internists, and Gynecologists. *Medical Care*, 22(1), 42-55.  
<https://doi.org/10.1097/00005650-198401000-00004>
- Schmidt, H. G., & Boshuizen, H. P. A. (1992). Encapsulation of Biomedical Knowledge. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 265-282). New York: Springer.
- Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53(2), 252-266.  
[https://doi.org/10.1016/0749-5978\(92\)90064-E](https://doi.org/10.1016/0749-5978(92)90064-E)
- Skovholt, T. M., Rønnestad, M. H., & Jennings, L. (1997). Searching for Expertise in Counseling, Psychotherapy, and Professional Psychology. *Educational Psychology Review*, 9(4), 361-369. <https://doi.org/10.1023/A:1024798723295>
- Spengler, P. M., & Pilipis, L. A. (2015). A Comprehensive Meta-Reanalysis of the Robustness of the Experience-Accuracy Effect in Clinical Judgment. *Journal of Counseling Psychology*, 62(3), 360-378. <https://doi.org/10.1037/cou0000065>
- Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . . Rush, J. D. (2009). The Meta-Analysis of Clinical Judgment Project: Effects of Experience on Judgment Accuracy. *The Counseling Psychologist*, 37(3), 350-399.  
<https://doi.org/https://doi.org/10.1177/0011000006295149>



- Tracey, T. J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in Psychotherapy: An Elusive Goal? *American Psychologist*, *69*(3), 218-229.  
<https://doi.org/10.1037/a0035099>
- Wagenaar, A. (2008). *Learning in internships: What and how students learn from experience* [Doctoral dissertation, Maastricht University]. National Academic Research and Collaborations Information System.
- Watkinson, J., Bristow, G., Auton, J., McMahon, C. M., & Wiggins, M. W. (2018). Postgraduate training in audiology improves clinicians' audiology-related cue utilisation. *International Journal of Audiology*, *57*(9), 681-687.  
<https://doi.org/10.1080/14992027.2018.1476782>
- Weiss, D. J., & Shanteau, J. (2003). Empirical Assessment of Expertise. *Human Factors*, *45*(1), 104-116. <https://doi.org/10.1518/hfes.45.1.104.27233>
- Wiggins, M., & O'Hare, D. (2003). Weatherwise: Evaluation of a Cue-Based Training Approach for the Recognition of Deteriorating Weather Conditions during Flight. *Human Factors*, *45*(2), 337-345. <https://doi.org/10.1518/hfes.45.2.337.27246>
- Wiggins, M. W. (2006). Cue-based processing and human performance. In W. Karwowski (Ed.), *Encyclopedia of Ergonomics and Human Factors* (2nd ed., pp. 641–645). Taylor & Francis.
- Wiggins, M. W. (2014). The role of cue utilisation and adaptive interface design in the management of skilled performance in operations control. *Theoretical Issues in Ergonomics Science*, *15*(3), 283-292. <https://doi.org/10.1080/1463922X.2012.724725>
- Wiggins, M. W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014a). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. *Safety Science*, *65*, 118-124. <https://doi.org/10.1016/j.ssci.2014.01.006>

- Wiggins, M. W., & Bollwerk, S. (2006). Heuristic-Based Information Acquisition and Decision Making Among Pilots. *Human Factors, 48*(4), 734-746.  
<https://doi.org/10.1518/001872006779166325>
- Wiggins, M. W., Brouwers, S., Davies, J., & Loveday, T. (2014b). Trait-based cue utilization and initial skill acquisition: Implications for models of the progression to expertise. *Frontiers in Psychology, 5*, 541-541. <https://doi.org/10.3389/fpsyg.2014.00541>
- Wiggins, M. W., Loveday, T., & Auton, J. (2015). *EXPERT Intensive Skills Evaluation (EXPERTise 2.0) Test*. [Software shell]. Macquarie University.
- Wigton, R. S. (1996). Social Judgement Theory and Medical Judgement. *Thinking & Reasoning, 2*(2), 175-190. <https://doi.org/10.1080/135467896394492>
- Wigton, R. S., Hoellerich, V. L., & Patil, K. D. (1986). How Physicians Use Clinical Information in Diagnosing Pulmonary Embolism: An Application of Conjoint Analysis. *Medical Decision Making, 6*(1), 2-11. <https://doi.org/10.1177/0272989X8600600102>
- Witteman, C. L. M., & Tollenaar, M. S. (2012). Remembering and diagnosing clients: Does experience matter? *Memory, 20*(3), 266-276.  
<https://doi.org/10.1080/09658211.2012.654799>
- Witteman, C. L. M., & Van Den Bercken, J. H. L. (2007). Intermediate effects in psychodiagnostic classification. *European Journal of Psychological Assessment, 23*(1), 56-61. <https://doi.org/https://doi.org/10.1027/1015-5759.23.1.56>
- Witteman, C. L. M., Weiss, D. J., & Metzmacher, M. (2012). Assessing diagnostic expertise of counselors using the Cochran-Weiss-Shanteau (CWS) Index. *Journal of Counseling and Development, 90*(1), 30-34. <https://doi.org/https://doi.org/10.1111/j.1556-6676.2012.00005.x>
- Yang, H., & Thompson, C. (2016). Capturing judgement strategies in risk assessments with improved quality of clinical information: How nurses' strategies differ from the

ecological model. *BMC Medical Informatics and Decision Making*, 16(7), 1-8.

<https://doi.org/https://doi.org/10.1186/s12911-016-0243-1>

**Appendix A: Study Flyer****Participants Needed**

For study exploring the expertise of clinical psychologists.

Are you enrolled in a Master of Psychology (Clinical) program or a registered clinical psychologist?

**Are you interested in testing your diagnostic accuracy?**

You are invited to participate via zoom or in person at the University of Adelaide. Both zoom and in person sessions will take place Tuesdays, Thursdays and Fridays from 9am to 5pm. Sessions will take approximately 15-25 minutes.

As compensation, you will go into the draw to win a \$100 Myer/Coles gift voucher.

To participate please email XXXX at [XXXX](mailto:XXXX) with the best three times that suit you from the sessions listed below.

9:00-9:30am  
9:30-10:00am  
10:00-10:30am  
10:30-11:00am  
11:30-12:00pm  
12:00-12.30pm  
12:30-1:00pm  
1:00-1:30pm  
1:30-2:00pm  
2:00-2:30pm  
2:30-3:00pm  
3:00-3:30pm  
3:30-4:00pm  
4:30-5:00pm

This data is being collected as a part of an Honours project under the supervision of XXXX and XXXX.

## **Appendix B: SONA Post**

### **Example Post on SONA**

#### **Title**

Are you better at diagnosing mental health conditions than clinical psychologists?

#### **Time**

35-45 minutes

#### **Summary**

In this experiment, you will be presented eight mental health case studies and you will be prompted to provide a primary and secondary diagnosis for each. You will rate which features present in the case studies assisted you in choosing a between a primary and secondary diagnosis. Finally, you will rate the importance of each feature in reaching the primary diagnosis.

#### **Eligibility**

You are eligible to participate in this study if you are a first-year Psychology student currently enrolled in Psychology 1A

#### **Compensation**

You will receive a full study credit for participation.

## Appendix C: Information Sheet

### Honours Project Information Sheet

#### Background

The purpose of this study is to investigate the expertise of clinical psychologists. Specifically, how the interpretation and use of clinical features changes with experience. In the broader expertise research, perceptual experts have been found to rapidly and accurately perceive the structural regularities that define categories and identities within their domain. Experts can identify significant features and their relations more efficiently than novices. For example, in a fingerprint study experts and novices differed in the features they chose, and experts tended to agree more with each other (Robson et al., 2019).

#### Eligibility

You are eligible to participate in this study if you are either:

- A first-year Psychology student currently enrolled in Psychology 1A or 1B
- Enrolled in a Masters of Psychology (Clinical) program
- A qualified and certified clinical psychologist

#### Task

In this experiment, you will be sequentially presented with eight mental health case studies on a laptop screen. Following each case study, you will be prompted to choose a primary and secondary diagnosis from the four options provided and rate your confidence in your primary diagnosis. You will rate each feature individually on a scale from -10 (supported secondary diagnosis) to +10 (supported primary diagnosis), with 0 indicating the feature equally supported both diagnoses. Finally, you will rate the importance of each feature in reaching the primary diagnosis.

#### Anticipated Time

The task should take approximately 35 to 45 minutes.

#### Associated Risks

This study itself is of a low-risk nature as there is no foreseeable risk of harm. However, in the case of emotional distress due to participation contact information for mental health services are provided.

Student Life Counselling Support (by appointment)

Ph: +61 8 8313 5663

Email: [counselling.centre@adelaide.edu.au](mailto:counselling.centre@adelaide.edu.au)

Adelaide University Crisis Line

Ph: 1300 167 654

Text: +61 488 884 107

Mental Health Triage

Ph: 131465

Since the study is taking place during the COVID-19 pandemic several health and safety protocols have been put in place:

- All surfaces and computer equipment will be sprayed and wiped down with disinfectant before and after each participant completes the task.

- Participants will be required to provide contact information (name, address, telephone number and email address) in the case of potential exposure to COVID-19.
- Participants will be asked to cancel their session if they are experiencing any cold or flu symptoms.

**Benefits**

The benefits of participating in this research include gaining course credit or going into a draw to win a \$100 gift voucher.

**Right to Withdraw**

If you wish to withdraw from the study at any point in time, you can do so without giving a reason.

**Data/Personal Information**

If you consent to participate in this study, your data will be de-identified as consent forms and contact tracing information will be assigned a unique participant code. This code will be used to match your demographic information, contact tracing information and consent to your experimental performance data. Data analysis will be undertaken by the researchers on the de-identified data set and only principal researcher (XXXX) will be able to re-identify the data.

If you give your permission, by signing the consent document, the results of this study will be detailed in a thesis to satisfy the Honours program requirements. Participant data will not be referred to by any identifiable features during the reporting of research results.

All digital information and data will be securely stored on the University of Adelaide database. Hardcopies of the consent forms will be kept by the principal supervisor in a locked storage cabinet. Records and materials will be retained by the University of Adelaide for a minimum of 5 years. De-identified data may be made available to future researchers via an open-science repository, which is a digital platform that retains research data and makes it freely available to other researchers, or similar.

**Contact information**

When you have read this information, the researcher (XXXX) and the principal investigator will discuss this study with you and answer any questions you may have. If you would like to know more at any stage please do not hesitate to contact us via the details outlined below.

XXXX  
Lecturer  
Ph: XXXX  
Email: XXXX

XXXX  
Psychology Honours Student/Researcher  
Ph: XXXX  
Email: XXXX

The study has been approved by the School of Psychology Human Research Ethics Subcommittee at the University of Adelaide (HREC-21/31). This research will be conducted according to the NHMRC National Statement on Ethical Conduct in Human Research (2007). For any questions about the ethical conduct of the research, please contact Professor Paul Delfabbro, Chair of the Human Research Ethics Subcommittee in the School of Psychology (paul.delfabbro@adelaide.edu.au).

**Appendix D: Consent Form****CONSENT FORM**

PROJECT: Does diagnostic reasoning in clinical psychology depend on feature use?

RESEARCHER: XXXX, School of Psychology, University of Adelaide

SUPERVISOR: XXXX, School of Psychology, University of Adelaide

LOCATION: Hughes Building, North Terrace

1. I agree to participate in the project named above, which is for research purposes. The particulars of the project, including details of the tasks, have been explained to me and provided to me via the participant information sheet.
2. I consent to any data gathered from this participation to be used for research purposes and to the data being stored in an online public repository (e.g., Open Science Framework).
3. I acknowledge that:
  - (a) the project is for the purpose of research;
  - (b) I have been informed that my involvement is voluntary and that I am free to withdraw from the project at any time without explanation or prejudice and to withdraw any unprocessed data previously supplied;
  - (c) the possible effects of the tasks have been explained to me to my satisfaction; and
  - (d) I have been informed that the confidentiality of the information I provide will be protected subject to any legal requirements.
4. I understand that:
  - (a) my real name will not be used in any publications arising from the research without my consent; and
  - (b) my participation in the research will have no effect on my academic grades, enrolment or future employment.

Name of participant:

---

Signature:

---

Date:

---

Please provide the following information for the purposes of COVID-19 contact tracing:

Telephone number:

Email:



**Appendix E: COVID-19 Screening Questions****COVID-19 Screening Questions**

1. Are you or anyone living in your household feeling unwell with any cold or flu-like symptoms such as cough, sore throat, headache, fatigue, body aches or a loss of taste and smell?
2. Have you felt feverish, had night sweats, or had a high temperature recorded recently?
3. Have you or anyone in your household returned from overseas or interstate in the last two weeks?
4. Have you or anyone in your household been in contact with a suspected or confirmed COVID-19 case in the last two weeks?
5. Have you been advised to be in home isolation for COVID-19?

## Appendix F: Case Studies

### Key

Shared feature

Unique feature

Contextual feature

### Case Study 1 – Male version

Charlie is a second-year university student who lives at home with his family. Ever since starting university last year Charlie has experienced a decrease in energy, which has led him to nap daily after returning home from university. Charlie has also had a significant decline in mood and even hanging out with friends or doing old hobbies does not bring him satisfaction anymore. He now finds his high school friends boring and thinks of them as less intelligent. Charlie has recently been obsessed with the idea that climate change is a conspiracy created by China to make Australian and US manufacturing non-competitive. He has joined an online forum where he posts frequently about this conspiracy, which has made him nervous that someone from the Chinese government will come looking for him. As the semester has progressed, his grades have suffered due to trouble focusing on readings or lectures. Charlie's dad thinks he is "just going through a phase", while his mum is worried about the noticeable decline in his mood and personal hygiene.

Primary diagnosis: Prodromal stage schizophrenia

Secondary diagnosis: Depression

### Case Study 2 – Female version

Alex is an 8-year-old girl who is brought in to see you because her mother is concerned about her seemingly random episodes of irritability and sadness. At school Alex struggles to make friends as she appears more concerned with her own internal world than playing with other children her age. Recently, while picking Alex up from school the teacher informed her mother that Alex was pushed over by another student and grazed her knee. Although Alex was visibly distressed and in pain she ignored her mother's attempt to hug or comfort her. In fact, Alex almost always avoids physical contact with her mother, relatives and peers. During your interactions with Alex you note her limited interest in therapy.

Primary diagnosis: Reactive attachment disorder

Secondary diagnosis: Autism spectrum disorder

### Case Study 3 – Male version

While driving to work Riley was hit by another car at a busy intersection. Riley sustained head injuries and broke his collarbone but received medical care early and is slowly recovering. Previously Riley had no history of mental health issues, yet in the 2 weeks since the crash he has experienced severe anxiety and frequently feels as if he can't breathe. Although Riley has always had difficulty falling asleep, he has found that he can now only manage to sleep 3 hours

on average a night. Riley sometimes feels disconnected and not in control of his body, as if he is floating in the air above it. Even after returning home from hospital Riley finds his home unfamiliar as his surroundings appear distorted and artificial like a movie. Riley has begun experiencing periods where he loses track of his current surroundings and feels as if he is trapped in the accident again.

Primary diagnosis: Acute stress disorder

Secondary diagnosis: Post-traumatic stress disorder

#### Case Study 4 – Female version

Taylor is a part-time retail worker and her partner works full-time as a nurse. In the past two months Taylor has experienced low self-esteem and her weight has steadily increased. Taylor has also been plagued by feelings of self-disgust and hatred. Taylor tends to eat a regular amount during the day, but at night while her partner is at work she experiences low mood and tends to eat palatable food to cope. In the moment, Taylor feels numb and cannot control her intake. However, once she finishes eating Taylor becomes extremely upset and consumed by guilt. These negative feelings after eating can sometimes cause Taylor to self-harm.

Primary diagnosis: Binge eating disorder

Secondary diagnosis: Depression

#### Case Study 5 – Male version

Kennedy is a 14-year-old boy who has been struggling at school. Kennedy's teacher notifies his parents that he has hurt several other students both during arguments and seemingly unprompted. This follows a recent incident where Kennedy stole money from the teacher's bag and lied about it. The teacher notes that Kennedy frequently refuses to follow instructions in the classroom and disrupts other students' learning. Kennedy has historically had difficulty making friends since he was a small child as he always wanted others to play his games and would yell at them if they did not. Kennedy's parents admit that their son argues with them every day, even about small things. The family dynamic has recently taken a turn as Kennedy kicked the family dog after an argument with his parents.

Primary diagnosis: Conduct disorder

Secondary diagnosis: Oppositional defiant disorder

#### Case Study 6 – Female version

Jordan is 24-years-old and lives at home with her parents. Originally Jordan sought help from her family GP for her headaches and gastrointestinal issues, but the GP decided to refer her to you. When you first meet with Jordan you notice that she speaks softly and avoids all eye contact. Jordan had poor school attendance and eventually dropped out of high school because she felt nauseous and panicked around other students and teachers. Jordan was always concerned that others were assessing her behaviour and thought she was strange. This has continued as Jordan has gotten older and has extended to other forms of social interaction, for example Jordan is afraid to make phone calls so she asks her sister to make them. While

Jordan desires a romantic relationship, she finds herself unwilling to get involved with other people due a preoccupation with being criticized or rejected.

Primary diagnosis: Avoidant personality disorder

Secondary diagnosis: Social anxiety disorder

### Case Study 7 – Male version

Denver is an 18-year-old who lives in a share house. This year he started university but changed his major four times before deciding to drop out. While in university Denver would frequently attend parties and binge drink which made it difficult for him to keep up with work and attend classes. At these parties, Denver would engage in unprotected sex with different partners but this has recently stopped as he began his first relationship. Within a week of meeting, Denver told the person he loved them and impulsively bought them expensive gifts. Denver has previously had issues maintaining friendships and will sometimes call his friends in desperation to make sure they still care about his. Denver also has a history of suicide attempts: once as a teen and in the last six months since beginning university.

Primary diagnosis: Borderline personality disorder

Secondary diagnosis: Bipolar disorder

### Case Study 8 – Female version

Harper is a 11-year-old girl who is home schooled. Since Harper was 8 years old, her parents noticed that she would tap and touch objects, like the dinner table, multiple times during lessons. Harper's mother also observed that she tends to shrug and sniff frequently during lessons. Recently, Harper has begun to repeat heard words or phrases which her mother has noted appear to repeat 3, 5 or 8 times. When Harper is asked about this behaviour she explains that it feels like needing to sneeze and although she can hold it off for a couple of minutes, it eventually happens. Harper has also shared that the symptoms seem less intense during the summer holidays, especially when she keeps busy with physical activities. However, when Harper is exhausted or anxious the symptoms worsen.

Primary diagnosis: Tic disorder

Secondary diagnosis: Obsessive compulsive disorder

## Appendix G: Regression Analyses R Script

# Data Analysis Notebook: Diagnostic Reasoning and Feature Use in Clinical Psychology 22/06/2021

## Libraries

```
# if(!"tinytex" %in% rownames(installed.packages())) install.packages("tinytex")
# tinytex::install_tinytex()
# # Install devtools package if necessary
# if(!"devtools" %in% rownames(installed.packages())) install.packages("devtools")
# # Install the stable development versions from GitHub
# devtools::install_github("crsh/papaja")
library(tidyverse)
library(ggplot2)
library(viridis)
library(RColorBrewer)
library(ez)
library(ggpubr)
library(rstatix)
library(broom)
library(corrplot)
source("dataSummary.R")
source("cohens_d.R")
source("https://bit.ly/2q4XQ66") ## geom_flat_violin()
```

## Data

```
## Import whole dataset
mydata <- read.csv("data.csv", header = TRUE)
mydata$Group <- factor(mydata$Group, c("Psychology Undergraduate Student", "Psychology Masters Student", "Clinical Psychologist"))
mydata$ID <- factor(mydata$ID)
```

## Demographics

### Compute descriptive stats for demographic variables...

```
## Age range
age <- mydata %>%
  group_by(Group) %>%
  summarise(Min = min(Age),
age

## Gender
gender <- mydata %>%
```

```

    group_by(Group, Gender) %>%
    count(Gender, sort = TRUE)
gender

# Education

education <- mydata %>%
  group_by(Group) %>%
  summarise(Min = min(Education_Numeric),
            Max = max(Education_Numeric),
            Median = median(Education_Numeric))
education

## Clinical Experience

experience <- mydata %>%
  filter(Group == "Clinical Psychologist") %>%
  summarise(Min = min(Experience_Numeric),
            Max = max(Experience_Numeric),
            Median = median(Experience_Numeric))
experience

```

## Correlations

```

mydata %>%
  select(Accuracy_1, Age, Education_Numeric, Experience_Numeric,
         Mean_Global_Feature_Weighting, Unique_Feature_Discriminability, Diagnostic_Discriminability_Unique,
         Diagnostic_Discriminability_Shared, Diagnostic_Discriminability_Contextual, Accuracy_2) -> Correlations

Correlations %>% cor() -> cor.matrix
cor.matrix

corrplot(cor.matrix, method="pie")

```

### Which measures of feature use are correlated with accuracy?

```

cor.test(Correlations$Accuracy_1, Correlations$Unique_Feature_Discriminability,
method = c("pearson"))

cor.test(Correlations$Accuracy_1, Correlations$Diagnostic_Discriminability_Unique,
method = c("pearson"))

cor.test(Correlations$Accuracy_1, Correlations$Diagnostic_Discriminability_Shared,
method = c("pearson"))

cor.test(Correlations$Accuracy_1, Correlations$Diagnostic_Discriminability_Contextual,
method = c("pearson"))

cor.test(Correlations$Unique_Feature_Discriminability, Correlations$Education_Numeric,
method = c("pearson"))

cor.test(Correlations$Unique_Feature_Discriminability, Correlations$Experience_Numeric,
method = c("pearson"))

cor.test(Correlations$Diagnostic_Discriminability_Unique, Correlations$Education_Numeric,
method = c("pearson"))

cor.test(Correlations$Diagnostic_Discriminability_Unique, Correlations$Experience_Numeric,
method = c("pearson"))

cor.test(Correlations$Diagnostic_Discriminability_Shared, Correlations$Education_Numeric,
method = c("pearson"))

```

```
cor.test(Correlations$Diagnostic_Discriminability_Shared, Correlations$Experience_N
  umeric, method = c("pearson"))

cor.test(Correlations$Diagnostic_Discriminability_Contextual, Correlations$Educatio
  n_Numeric, method = c("pearson"))

cor.test(Correlations$Diagnostic_Discriminability_Contextual, Correlations$Experien
  ce_Numeric, method = c("pearson"))
```

### Does education and/or clinical experience predict accuracy?

```
cor.test(Correlations$Accuracy_1, Correlations$Education_Numeric, method =
  c("pearson"))

cor.test(Correlations$Accuracy_1, Correlations$Experience_Numeric, method =
  c("pearson"))

cor.test(Correlations$Accuracy_2, Correlations$Education_Numeric, method =
  c("pearson"))

cor.test(Correlations$Accuracy_2, Correlations$Experience_Numeric, method =
  c("pearson"))
```

## Assumptions

```
## Linearity Assumption
ggscatter(
  mydata, x = "Unique_Feature_Discriminability", y = "Accuracy_1",
  facet.by = "Group",
  short.panel.labs = FALSE
)+
  stat_smooth(method = "loess", span = 0.9)

## Homogeneity of Regression Slopes Assumption
mydata %>%
  anova_test(
    Accuracy_1 ~ Unique_Feature_Discriminability + Group
  Unique_Feature_Discriminability*Group)

## Normality of Residuals Assumption

model <- lm(Accuracy_1 ~ Unique_Feature_Discriminability + Group, data = mydata) #
  Fit model
model.metrics <- augment(model) %>%
  select(-.hat, -.sigma, -.fitted)
head(model.metrics, 3) # Inspect model
```

## Inferential Analysis: GLM Approach

### Relationship between education, clinical experience, and discrimination of unique vs. other features...

```
glm_1 <- glm(Unique_Feature_Discriminability ~ Education_Numeric +
  Experience_Numeric, family=gaussian, data = mydata)
summary(glm_1)
```

### Relationship between education, clinical experience, and discriminant use of unique features in primary and secondary diagnoses...

```
glm_2 <- glm(Diagnostic_Discriminability_Unique ~ Education_Numeric +  
Experience_Numeric, family=gaussian, data = mydata)  
summary(glm_2)
```

**Relationship between education, clinical experience, and discriminant use of shared features in primary and secondary diagnoses...**

```
glm_3 <- glm(Diagnostic_Discriminability_Shared ~ Education_Numeric +  
Experience_Numeric, family=gaussian, data = mydata)  
summary(glm_3)
```

**Relationship between education, clinical experience, and discriminant use of contextual features in primary and secondary diagnoses...**

```
glm_4 <- glm(Diagnostic_Discriminability_Contextual ~ Education_Numeric +  
Experience_Numeric, family=gaussian, data = mydata)  
summary(glm_4)
```

**Relationship between education, clinical experience, and accuracy of primary diagnosis...**

```
glm_5 <- glm(Accuracy_1 ~ Education_Numeric + Experience_Numeric, family=gaussian,  
data = mydata)  
summary(glm_5)
```

**Education and unique feature discriminability as predictors of accuracy of primary diagnosis**

```
glm_6 <- glm(Accuracy_1 ~ Education_Numeric + Unique_Feature_Discriminability,  
family=gaussian, data = mydata)  
summary(glm_6)
```

```
glm_7 <- glm(Accuracy_1 ~ Experience_Numeric + Diagnostic_Discriminability_Contextu  
al, family=gaussian, data = mydata)  
summary(glm_7)
```

```
glm_8 <- glm(Accuracy_2 ~ Education_Numeric + Unique_Feature_Discriminability,  
family=gaussian, data = mydata)  
summary(glm_8)
```



## Appendix H: ANOVA/ANCOVA Analyses R Script

# Inferential Analysis: ANOVA/ANCOVA Approach

Effect of Group on discrimination of unique vs. other features...

```
ANOVA_1 <- mydata %>%
anova_test(Unique_Feature_Discriminability ~ Group)

## Coefficient covariances computed by hccm()
get_anova_table(ANOVA_1)
```

Effect	DFn	DFd	F	p	p<.05	ges
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>
1 Group	2	33	0.691	0.508		0.04
1 row						

```
options(contrasts=c("contr.treatment","contr.poly")) # Choose contrast options
contrasts <- lm(Accuracy_1 ~ Unique_Feature_Discriminability + Group, mydata) # Compute contrasts
summary(contrasts) # Show contrast results
```

```
## Call:
## lm(formula = Accuracy_1 ~ Unique_Feature_Discriminability + Group,
##     data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.31262 -0.11372 -0.01583  0.07657  0.41602
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.26990    0.05096   5.296  8.4e-06 ***
## Unique_Feature_Discriminability 0.17228    0.05541   3.109  0.00392 **
## GroupPsychology Masters Student 0.11812    0.13092   0.902  0.37369
## GroupClinical Psychologist     0.13375    0.06831   1.958  0.05901 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1779 on 32 degrees of freedom
## Multiple R-squared:  0.3448, Adjusted R-squared:  0.2834
## F-statistic: 5.613 on 3 and 32 DF, p-value: 0.003295
```

Effect of Group on use of unique features to discriminate primary and secondary diagnoses...

```
ANOVA_2 <- mydata %>%
anova_test(Diagnostic_Discriminability_Unique ~ Group)

## Coefficient covariances computed by hccm()
get_anova_table(ANOVA_2)
```

Effect	DFn	DFd	F	p p<.05	ges
<chr>	<dbl>	<dbl>	<dbl>	<dbl> <chr>	<dbl>
1 Group	2	33	1.56	0.225	0.086
1 row					

```
options(contrasts=c("contr.treatment","contr.poly")) # Choose contrast options
contrasts <- lm(Accuracy_1 ~ Diagnostic_Discriminability_Unique + Group, mydata) #
Compute contrasts117summary(contrasts) # Show contrast results
```

```
##
## Call:
## lm(formula = Accuracy_1 ~ Diagnostic_Discriminability_Unique +
##     Group, data = mydata)
##
## Residuals:
##     Min       1Q   Median       3Q      Max
## -0.33786 -0.10870 -0.02911  0.09065  0.41380
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.31097    0.04996   6.224 5.68e-07 ***
## Diagnostic_Discriminability_Unique  0.08619    0.03901   2.210  0.0344 *
## GroupPsychology Masters Student    0.06241    0.14149   0.441  0.6621
## GroupClinical Psychologist         0.13325    0.07374   1.807  0.0802
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1891 on 32 degrees of freedom
## Multiple R-squared:  0.2598, Adjusted R-squared:  0.1904
## F-statistic: 3.744 on 3 and 32 DF,  p-value: 0.02061
```

Effect of Group on use of shared features to discriminate primary and secondary diagnoses...

```
ANOVA_3 <- mydata %>%
anova_test(Diagnostic_Discriminability_Shared ~ Group)
## Coefficient covariances computed by hccm()
```

Effect	DFn	DFd	F	p p<.05	ges
<chr>	<dbl>	<dbl>	<dbl>	<dbl> <chr>	<dbl>
1 Group	2	33	0.697	0.505	0.041
1 row					

```
get_anova_table(ANOVA_3)
options(contrasts=c("contr.treatment","contr.poly")) # Choose contrast options
contrasts <- lm(Accuracy_1 ~ Diagnostic_Discriminability_Shared + Group, mydata) #
Compute contrasts117summary(contrasts) # Show contrast results
```

```
##
## Call:
## lm(formula = Accuracy_1 ~ Diagnostic_Discriminability_Shared +
##     Group, data = mydata)
##
## Residuals:
##     Min       1Q   Median       3Q      Max
## -0.43468 -0.13642  0.00231  0.11156  0.40336
##
```

```
## Coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.43898    0.05060   8.675 6.51e-10 ***
## Diagnostic_Discriminability_Shared -0.09759    0.05346  -1.825  0.0773 .
## GroupPsychology Masters Student    0.10850    0.14229   0.763  0.4513
## GroupClinical Psychologist         0.20157    0.07406   2.722  0.0104 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1932 on 32 degrees of freedom
## Multiple R-squared:  0.2273, Adjusted R-squared:  0.1549
## F-statistic: 3.138 on 3 and 32 DF,  p-value: 0.03883
```

Effect of Group on use of contextual features to discriminate primary and secondary diagnoses...

```
ANOVA_4 <- mydata %>%
anova_test(Diagnostic_Discriminability_Contextual ~ Group)
## Coefficient covariances computed by hccm()
get_anova_table(ANOVA_4)
```

<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<dbl>
1 Group	2	33	0.867	0.429		0.05
1 row						

```
options(contrasts=c("contr.treatment","contr.poly")) # Choose contrast options
contrasts <- lm(Accuracy_1 ~ Diagnostic_Discriminability_Contextual + Group,
mydata ) # Compute contrasts
summary(contrasts) # Show contrast results
```

```
##
## Call:
## lm(formula = Accuracy_1 ~ Diagnostic_Discriminability_Contextual +
##     Group, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38574 -0.10809 -0.04081  0.14990  0.36002
##
## Coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.38847    0.04392   8.845 4.18e-10 ***
## Diagnostic_Discriminability_Contextual -0.02627    0.05335  -0.492  0.6258
## GroupPsychology Masters Student    0.13533    0.15252   0.887  0.3815
## GroupClinical Psychologist         0.17536    0.07612   2.304  0.0279 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2022 on 32 degrees of freedom
## Multiple R-squared:  0.1533, Adjusted R-squared:  0.07389
## F-statistic: 1.931 on 3 and 32 DF,  p-value: 0.1444
```

Effect of Group on Accuracy of primary diagnosis (ANOVA) accounting for variation in unique feature discriminability (ANCOVA\_1)...

```
ANOVA_5 <- mydata %>%
anova_test(Accuracy_1 ~ Group)
## Coefficient covariances computed by hccm()
get_anova_table(ANOVA_5)
```

Effect	DFn	DFd	F	p p<.05	ges
<chr>	<dbl>	<dbl>	<dbl>	<dbl> <chr>	<dbl>
1 Group	2	33	2.84	0.073	0.147

```
options(contrasts=c("contr.treatment","contr.poly")) # Choose contrast options
contrasts_2 <- lm(Accuracy_1 ~ Group, mydata) # Compute contrasts
summary(contrasts_2) # Show contrast results
```

```
## Call:
## lm(formula = Accuracy_1 ~ Group, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41400 -0.09508 -0.04054  0.15700  0.33292
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.38108    0.04080   9.340 8.7e-11 ***
## GroupPsychology Masters Student  0.11892    0.14711   0.808  0.4247
## GroupClinical Psychologist      0.17592    0.07524   2.338  0.0256 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1999 on 33 degrees of freedom
## Multiple R-squared:  0.1469, Adjusted R-squared:  0.09515
## F-statistic:  2.84 on 2 and 33 DF,  p-value: 0.07275
```

# ANCOVA

```
ANCOVA_1 <- mydata %>%
anova_test(Accuracy_1 ~ Unique_Feature_Discriminability + Group)
```

```
## Coefficient covariances computed by hccm()
```

```
get_anova_table(ANCOVA_1)
```

Effect	...	...	F	p p<.05	ges
<chr>	<dbl>	<dbl>	<dbl>	<dbl> <chr>	<dbl>
1 Unique_Feature_Discriminability	1	32	9.667	0.004 *	0.232
2 Group	2	32	2.109	0.138	0.116
2 rows					

```
options(contrasts=c("contr.treatment","contr.poly")) # Choose contrast options
contrasts_2 <- lm(Accuracy_1 ~ Diagnostic_Discriminability_Contextual + Group, mydata) # Compute contrasts
summary(contrasts_2) # Show contrast results
```

```
##
## Call:
## lm(formula = Accuracy_1 ~ Diagnostic_Discriminability_Contextual +
##      Group, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38574 -0.10809 -0.04081  0.14990  0.36002
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.38847    0.04392   8.845 4.18e-10 ***
```

```
## Diagnostic_Discriminability_Contextual -0.02627    0.05335  -0.492    0.6258
```