Journal of Affective Disorders 227 (2018) 698–704

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Distinguishing transient versus stable aspects of depression in New Zealand Pacific Island children using Generalizability Theory

Janis Paterson[a], Oleg N. Medvedev[b], Alexander Sumich[a,c], El-Shadan Tautolo[a], Christian U. Krägeloh[a], Rose Sisk[d], Robert K. McNamara[e], Michael Berk[f], Ajit Narayanan[g], Richard J. Siegert[a,*]

[a] School of Public Health and Psychosocial Studies, Auckland University of Technology, Auckland, New Zealand
[b] Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand
[c] Division of Psychology, Nottingham Trent University, Nottingham, United Kingdom
[d] Biostatistics and Epidemiology, Auckland University of Technology, New Zealand
[e] Department of Psychiatry and Behavioral Neuroscience, University of Cincinnati College of Medicine, United States
[f] Deakin University, IMPACT Strategic Research Centre, School of Medicine, Geelong, Victoria, Australia
[g] School of Engineering, Computer and Mathematical Sciences, Auckland University of Technology, New Zealand

ARTICLE INFO

Keywords:
Depression
GeneralizabilityTheory
Children's Depression Inventory
State
Trait

ABSTRACT

Background: The distinction between temporary versus enduring or state/trait aspects of depression is important. More precise distinction would improve understanding of the aetiology of depression and those aspects most amenable to intervention thus identifying more homogeneous, dynamic targets for clinical trials. Generalizability Theory has been proposed as useful for disentangling state and trait components of psychopathology.

Methods: We applied Generalizability Theory to determine the relative contributions of temporary and enduring aspects of depression in a widely used screening measure of depression the – 10-item Children's Depression Inventory (CDI-10; Kovacs, 1985). Participants were children of Pacific Island descent living in New Zealand (n = 668). Data were collected at ages – 9, 11, and 14 years.

Results: The CDI-10 demonstrated acceptable generalizability across occasions (G = 0.79) with about one third of variance in total scores attributed to temporary and two thirds to more enduring aspects of depression. There were no other significant sources of error variance. Two items were identified as more sensitive than the remaining eight to more dynamic symptoms.

Limitations: Studies with briefer test-retest intervals are warranted. Use of this Pacific Island cohort limits generalizability of findings to other cultures and ethnicities. No data were collected on whether participants had received intervention for depression.

Conclusions: While the CDI-10 reliably measures both stable and transient aspects of depression in children, the scale does not permit clear distinction between them. We advocate application of Generalizability Theory for developing state/trait depression measures and determining which existing measures are most suitable for capturing modifiable features of depression.

1. Introduction

The distinction between dynamic and more stable aspects of behavior is well established in psychology (Chaplin et al., 1988). Generally, relatively stable or enduring aspects of an individual's behavior are referred to as a *trait*, while more dynamic aspects are referred to as a *state*, meaning changeable behavior that is much more context bound or situation specific. The distinction between such stable and dynamic aspects of behavior has sometimes been demonstrated by comparing test - retest correlations for the state and trait subscales of a measure, e.g., the State-Trait Anxiety Inventory (STAI; Gaudry et al., 1975), with a higher correlation expected for trait than state components. For example, Barnes et al. (2002) reviewed the reliability of the STAI and found a mean test-retest correlation of 0.88 (range 0.82–0.94) for trait anxiety and 0.70 (range 0.34–0.96) for state anxiety, based on seven studies reporting test-retest reliability. An obvious limitation of this

* Correspondence to: Auckland University of Technology, Psychology, School of Public Health and Psychosocial Studies, Room AR321, 90 Akoranga Drive, Northcote, Private Bag 92006, Auckland 1142, New Zealand.
E-mail address: richard.siegert@aut.ac.nz (R.J. Siegert).

https://doi.org/10.1016/j.jad.2017.11.075
Received 17 August 2017; Received in revised form 9 November 2017; Accepted 12 November 2017
Available online 16 November 2017
0165-0327/

approach to distinguishing state from trait is that it uses total score correlations and fails to examine the relative strengths of individual items.

An alternative approach to assessing state versus trait has been to use structural equation modelling to compare the covariance structures of a state/trait measure on at least two separate occasions (Cole et al., 2005). Using this "latent trait-state-occasion model" with the Beck Depression Inventory II (BDI-II) in two large samples of young people Wu reported a trait component accounting for more than 50% of variance and an 'occasion-specific factor' that explained between 7% and 12% of variance (Wu, 2016, p.39). LaGrange et al. (2011) applied the latent trait-state-occasion model to four measures of depressive cognitions and symptoms in a longitudinal study of 515 children and adolescents and found three of these measures characterized by "two types of longitudinal factors: a time invariant (or trait-like) factor and a series of time-varying (more state-like) factors" (p.13). Of particular relevance to the present study was their fourth measure - the Children's Depression Inventory - was almost entirely accounted for by time varying or state like variance. While this approach has the advantage that it can compare the strength of the loadings of individual items on state or trait factors at different points in time it fails to partial out the precise proportion of variance across occasions due to the person, the occasion and the item (as well as their interactions). A method which does exactly this, namely to account for all major sources of variance in a measurement situation, and arguably (Medvedev et al., 2017) should be the preferred technique for demonstrating a state-trait distinction is Generalizability Theory.

1.1. Generalizability Theory

Generalizability or G Theory was developed by Cronbach and represents a major advance upon Classical Test Theory (CTT) particularly in regard to evaluating reliability (Cronbach et al., 1963). While CTT conceptualizes test scores as the sum of a true score and error variance, G Theory uses analysis of variance (ANOVA) to calculate precise estimates for the error variance due to each important measurement *facet*, where facet refers to a distinct element that might influence variance and error in test scores in any testing situation. For example, the persons tested, the test items and the testing occasion are three examples of facets. Hence, whereas CTT restricts analysis of reliability and error variance to a single element such as the test items (i.e., Cronbach's alpha), the occasion (test-retest) or the rater (inter-rater), G Theory permits the researcher to break reliability down into all the important facets contributing to measurement error in a single analysis. In G terms the variance associated with participants or persons is considered the central concern and is known as the *differentiation facet* with other facets (e.g., items, occasion, rater) viewed as sources of measurement error.

An important step in a Generalizability analysis is the calculation of individual variance components for each facet (e.g., person, item, occasion) in the measurement situation. These variance components are first calculated by a conventional ANOVA and corrected by a method known as 'Whimbey's correction' which accounts for the type of sampling involved (i.e. random, fixed or random finite). This also includes calculation of a *G coefficient* which represents how generalizable the test scores are across different situations. Bloch and Norman (2012, p. 968) describe the G coefficient as the ratio of signal to noise or 'true' variance to 'true + error variance' and provide details on its derivation and calculation. In most Generalizability studies there are two stages known as the G-study and the D-study. The G-study involves a standard factorial ANOVA with the calculation of variance components for each facet and for their interactions (rather than the significance tests typically associated with an ANOVA). The D-study allows the researcher to then estimate the impact on reliability of variations in different facets such as increasing the number of participants or the number of items in a scale.

As an illustration, medical education is one setting where G theory has been widely used (e.g., Prion et al., 2016). For example, when nursing students are examined on clinical skills, they might encounter ten structured clinical problems (or *stations*) that they are required to respond to in the presence of an examiner who rates their performance on a 1–10 scale for each problem. Assuming a different examiner (i.e. rater) at each station, a G Theory approach permits us to calculate the variance component for each facet – namely student, station, and rater - and their interactions. Consider another example from a mental health setting. In examining the inter-rater reliability of a new rating scale for assessing the severity of depression in outpatients we might have the following facets: Person (the differentiation facet), rater, and occasion. The calculation of variance components for each facet and their interactions allows us to determine precisely how much variance is due to the patient (Person facet) and changes in their condition (Person × Occasion) as well as the reliability of raters. High inter-rater reliability will be reflected in a very small variance component for the facet Rater. In the present article we argue that G Theory may have particular value for distinguishing between state and trait or stable versus enduring aspects of behavior and depression in particular.

1.2. Stable versus transient features of depression

A key issue for measuring change in symptoms of psychopathology is the extent to which the underlying construct is stable or changes over time – the state-trait distinction. This distinction between stable and dynamic symptoms has been applied less often for conceptualizing depression than for anxiety. However, there has been increasing interest recently in separating dynamic versus more stable aspects of depression from both neurobiological and personality perspectives (Bhagwagar et al., 2002; Graham et al., 2013; Natoli et al., 2016; Wu, 2016). In a recent psychometric study, Wu (2016) tested a latent state-trait-occasion model using the Beck Depression Inventory II (BDI-II) in two cohorts, of adolescents and young adults, concluding that the BDI-II measures "both trait-like and occasion-specific variances for individuals during late adolescence to early adulthood" (p.48). Lakes and Hoyt (2009) have argued that G Theory is a particularly useful conceptual framework for measuring the stable and transient components of a latent construct among adolescents where development is paramount. A recent paper by Chavez et al. (2016) demonstrated this approach with the Adolescent Quality of Life-Mental Health Scale (AQOL-MHS) in adolescents attending mental health clinics assessed on three occasions over eight months. The focus of that study was on the reliability of change scores and the authors used a method developed by Cranford et al. (2006) based on G Theory to compare conventional estimates of reliability (Cronbach's alpha, test-retest) with GT estimates. The authors observed that the reliability of change assessed by G Theory for the three subscales of the AQOL-MHS was adequate but consistently lower than Cronbach's alpha for each time point. They concluded Cronbach's alpha is sufficient when comparing scores across people at the same point in time but the reliability of change scores must be considered when comparing the change in persons across multiple time points.

1.3. The Children's Depression Inventory

The Children's Depression Inventory is a widely used measure of depression in children and adolescents. It is a 27-item self-report questionnaire for quantifying the severity of symptoms of depression in young people aged 7–17 years. We are aware of only one previous article using G Theory to investigate the psychometric characteristics of the CDI (Crowley et al., 1994). In that study they administered the full 27-item CDI twice, with a 28-week retest interval, to 164 children aged 11–16 years (mean age 12.6) in Texas schools. Crowley et al. (1994) found a relatively low generalizability coefficient for a single CDI administration (0.63) compared with repeat testing (0.81) and advocated multiple assessments before diagnosing depression in a child. They