An Empirical Evaluation of the Reversal Theory State Measure Using Three Running Brand Video Commercials

Leo Watkins, Marcia Wilson, Mark McDermott, and Richard Buscombe
University of East London

This paper reports on the first known empirical use of the Reversal Theory State Measure (RTSM) since its publication by Desselles et al. (2014). The RTSM was employed to track responses to three purposely-selected video commercials in a between-subjects design. Results of the study provide empirical support for the central conceptual premise of reversal theory, the experience of metamotivational reversals and the ability of the RTSM to capture them. The RTSM was also found to be psychometrically sound after adjustments were made to two of its three component subscales. Detailed account and rationale is provided for the analytical process of assessing the psychometric robustness of the RTSM, with a number of techniques and interpretations relating to component structure and reliability discussed. Agreeability and critique of the two available versions of the RTSM – the bundled and the branched – is also examined. Researchers are encouraged to assist development of the RTSM through further use, taking into account the analysis and recommendations presented.

Keywords: State measure, metamotivational state, metamotivational reversals, reversal theory

The Reversal Theory State Measure (RTSM) represents the first fully comprehensive state measure to be created for the measurement of the eight metamotivational states proposed by reversal theory. The RTSM has been developed over several years (Desselles & Murphy, 2011; Desselles, Murphy, & Theys, 2014; Young, Desselles, Lee, & Apter, 2005) culminating in open access for researchers to two versions of the instrument. The emergence of the RTSM represents an important development within reversal theory, where previous state measures have failed to account for the full range of experiences posited by the theory.

Previous measures have only measured a single subset of reversal theory states: the Telic State Measure (TSM; Svebak & Murgatroyd, 1985), the Negativism State Measure (O'Connor, 1992), the Somatic State Questionnaire (SSQ; Cook, Gerkovich, Potocky, & O'Connell, 1993), the Telic/Paratelic State Instrument (T/PSI; Calhoun, 1995; O'Connell & Calhoun, 2001), the Autic Mastery-Sympathy State Measure (O'Connell & Brooks, 1997). Other important empirical measures within the theory have tended to be trait rather than state based such as The Motivational Style Profile (MSP; Apter, Mallows, & Williams, 1998) or the Negativism Dominance Scale (McDermott, 1988).

Correspondence concerning this article should be addressed to Leo Watkins, School of Health, Sport and Bioscience, University of East London, Water Lane, Stratford, E15 4LZ. E-mail: l.watkins@uel.ac.uk

The RTSM is able to measure and classify metamotivational states in two ways: through both "branched" and "bundled" versions of the instrument. The more extensive "branched" version works by presenting respondents with a fixed-choice format (one or the other) of three opposing statements for each pair of states within the means-ends, rules, and relationship-orientated domains. Respondents must chose one from each pair of opposing statements to best reflect how they currently feel. Respondents must also then score the selected statement on an interval scale of 1 (weaker) to 3 (stronger). The prevailing metamotivational state is then calculated by summing the scores for each of the three pairs of statements within each domain. In the event of a tie, the state with more responses is deemed most salient. Thus a participant scoring telic (2) on the first item but scoring paratelic (1) on the second and third items within the telic-paratelic subscale would be classified as paratelic.

Once the initial set of nine items has been answered, respondents then have a further three items to complete. These are based on "crossed" transactional pairs from either the autic and mastery/sympathy states or the alloic and mastery/sympathy states. Which set of crossed pairs respondents are directed to is based on their answers to the items comprising the autic and alloic states. Thus for example, once a respondent has indicated they are in the autic rather than alloic state it must then be established whether they are in the autic-mastery or autic-sympathy state. Successful completion of the measure will involve respondents completing a total of twelve of the fifteen items presented and classification across all four metamotivational domains.

In contrast to the longer and interval based scale of the branched RTSM, the bundled version works by presenting the same statements for each metamotivational state in a single choice, categorical manner. Instead of making three separate choices from pairs of opposing statements to determine the salient metamotivational state, respondents make a choice of one of two "bundles" of opposing statements. For example, a conformist "bundle" consisting of the statements (with the prefix "I wanted to"): Do what I'm supposed to do. Do what's expected of me. Do my duty. This process is carried out separately for the means-ends and rules orientated domains and is then followed by a choice of four bundles of statements reflecting the four crossed transactional states. Thus using the bundled version of the RTSM requires the completion of only three items.

Theoretical Underpinning of the Two Versions

In outlining the development of the RTSM, Desselles et al. (2014) identify the principal merits of each of the two versions. In addition to the obvious advantage of speed of completion and scoring, the authors point out that the bundled version carries two further benefits over the branched RTSM. Firstly, by presenting each metamotivational state using a combination of three statements, the bundled RTSM can be seen to provide a more thorough description of each state, clarifying for the respondent what is being measured. Secondly, by forcing a single choice between one state and another, it provides a clearer affirmation of an individual's metamotivational state than that comparing scores across two opposing states. This also better supports reversal theory's theoretical assertion of bi-stability; an individual is simply either in one state or another rather than being slightly more in one state than another.

However, in creating these advantages in efficiency and simplicity over the more complex branched version, the use of the bundled RTSM involves a significant trade off in the type of data it provides. In employing a dichotomous scoring method, the bundled RTSM provides categorical data, with the required non-parametric tests providing less analytical flexibility than their parametric counterparts. In addition Desselles et al. (2014) have also suggested that the combining of descriptive statements within the bundled version may cause confusion in some respondents.

In contrast, the branched RTSM carries the twin advantages of presenting respondents with a simple choice between two statements each time and the subsequent production of more versatile, parametric data. The Branched version of the RTSM can also claim to have investigated respondents' metamotivational state more thoroughly as it is in effect able to test each domain on three separate occasions. The limitation to the way the branched RTSM is structured however, is that the advanced branching logic would require

the dynamic programming of an online instrument to function exactly as intended.

These differences in the two versions of the RTSM articulate an important objective in the development of the measure and in the study as carried out here; to assess whether they produce the same results when classifying participants. In addition to addressing this question, the current study presents an opportunity to test the psychometric robustness of the RTSM in an intervention-based study where states have been intentionally manipulated. The component structure of an instrument may vary from sample to sample (Hinkin 1995) therefore it is also valuable to examine the RTSM's structure on a different sample to that used by its authors.

Method

Design and Purpose

The current study was designed with the primary objective of measuring changes in metamotivational state in response to watching one of three purposely-selected video commercials. The commercials selected for the study were chosen following a rigorous reversal theory analysis of their content in a prior study. The current study intended to examine whether identified differences in the metamotivational states represented in the commercials resulted in differences in the states experienced by each video's audience. However, the focus of what is presented here is not to discuss any differences resulting as a function of each video but rather make rigorous examination of the instruments used.

Sample

After being piloted on a small sample of both undergraduate and postgraduate students, the questionnaire survey was administered to a sample of 501 undergraduate students at a London university. The demographic profile of the sample was as follows; M age was 22.1 years ($SD = 4.4$), with 43.7% males and 56.3% females. The sample was 36.9% White, 26.3% Asian, 23.9% Black, 9% mixed ethnicity, 2.3% Chinese or other and the remaining 1.9% undeclared.

Materials

In addition to the outlined branched and bundled versions of the RTSM, three different video commercials were used to form the visual intervention component of the study. The commercials were purposely selected from leading brands in the running shoe market in which the content and duration of each video commercial was comparable. The three brands used were ASICS, Saucony, and Mizuno with the most recently available commercial video relating specifically to running at the time of the study's conception used for each brand. The brands chosen were particularly selected for their primary market position as running brands rather

than general or casual athletic brands. Each of the videos used was approximately one minute long with a clear focus on running. As much as was possible, each video was shown to approximately one third of the sample.

Procedure

Six separate groups of participants were recruited for the study. One of the three videos was shown to each group on a projector screen housed within in a lecture theatre setting. The three videos were split across the six participant groups, so that each video was shown to two different groups, with no group being shown more than one of the videos. On each of the six occasions data was collected, each participant involved in the study had clear and unobscured audio-visual access to the video material provided.

The study was administered using a pen and paper version of the RTSM. Participants were instructed to stop part way through to be shown one of three commercials before completing the questionnaire. As directed by the RTSM's authors, the process of randomizing items was followed to determine how they were displayed on the questionnaire. Since the questionnaire was unable to utilize dynamic programming, it was not possible to enforce the "adaptive questioning" (Desselles et al. 2014, p. 16) of the branched version's crossed transactional items. Therefore, participants were left to complete all 15 items with the irrelevant crossed transactional items then discarded upon data entry.

Participants completed the bundled RTSM in order to document their pre-video metamotivational state. They were then shown one of the three commercials. To measure any metamotivational changes, we asked the participants to then complete bundled RTSM a second time, documenting their post-viedo metamotivational state. Prior evaluation by Desselles et al. (2014) suggested several advantages of the bundled RTSM over the branched version in this situation, including concerns about the length of the branched RTSM. However, after wathcing the video participants did also complete the branched version of the RTSM (prior to the bundled). This was to enable assessment of how accurately the two versions documented the same metamotivational states.

Results

Overview of Recorded States and Reversals

Using the bundled version of the RTSM as the post-intervention measure, reversals were recorded across 14 of the 16 possible reversals (see Table 1). In total, 222 reversals were documented during the study, confirming both the phenomenological basis of reversal theory and the ability of the bundled RTSM to capture the phenomenology of reversals. The stimulus for these reversals and the extent to which they can be attributed to the intervention materials used presents

Table 1

Metamotivational reversals documented using the bundled RTSM as a repeated measure.

Reversal Type: Pre-Post Measure	N (Total Pre-State)	% of Respondents from Pre-State
Telic-Paratelic	34 (293)	11.6
Paratelic-Telic	43 (93)	46.2
Conformist-Negativist	33 (359)	9.2
Negativist-Conformist	12 (29)	41.4
Autic Mastery-Alloic Mastery	4 (108)	3.7
Autic Mastery-Autic Sympathy	3 (108)	2.8
Autic Mastery-Alloic Sympathy	0 (108)	0
Autic Sympathy-Autic Mastery	13 (20)	65
Autic Sympathy-Alloic Mastery	2 (20)	10
Autic Sympathy-Alloic Sympathy	0 (20)	0
Alloic Mastery-Autic Mastery	29 (56)	51.8
Alloic Mastery-Autic Sympathy	2 (56)	3.6
Alloic Mastery-Alloic Sympathy	3 (56)	5.4
Alloic Sympathy-Autic Mastery	33 (84)	39.3
Alloic Sympathy-Autic Sympathy	10 (84)	11.9
Alloic Sympathy-Alloic Mastery	1 (84)	1.2

a key question for the study. However, it must be reiterated that this is not the focus of this particular paper.

Principal Components

A Principal Components Analysis (PCA) was conducted in order to reduce and simplify the RTSM's variables into their primary descriptors (Jolliffe 2002). PCA for the current study was carried out on the first nine items of the branched RTSM comprising the telic-paratelic, conformist-negativist and autic-alloic pairs. It was not possible to include the alloic-mastery/alloic-sympathy pair due to insufficient cell size. Data was entered into SPSS (version 20) and cleaned of missing cases, reducing the sample to 395, still adequate in size to meet he criteria for performing PCA (e.g. Comfrey & Lee, 1992; Nunnally, 1978).

Principal components were identified using established criteria based on examination of eigenvalues (Kaiser 1970) and the scree plot (Cattell 1966). Rotation of identified factors was performed orthogonally since the theoretical structure of the RTSM implies that underlying factors would be uncorrelated. Varimax rotation was used in order to maximize variance of loadings between components (Tabachnick & Fidell, 2013). Retained factors were then subjected to a reliability analysis of their internal consistency.

Results of the initial PCA indicated there to be three principal components (see Tables 2 and 3), supporting the intended structure of the uncrossed pairs within the branched RTSM. However, this finding was not reflected satisfactorily across all factor-identifying methods used, leaving us to in-

Table 2
Rotated Factor Loadings for the Uncrossed pairs of the Branched RTSM – Initial Three-Factor Solution.

| | | Principal Components | | |
| | | Factor Loadings | | |
Pair	Item Anchor	1	2	3
TP 3	Accomplish something for the future/Enjoy myself at this moment	**.766**	.026	.048
TP 2	Do something serious/Do something playful	**.709**	.132	-.174
TP 1	Do something crucial/Do something of no great concern	**.707**	.135	.010
AA 1	Focus on my own needs/Focus on others' needs	.205	**.819**	-.065
AA 3	Do something for myself/Do something for others	.298	**.713**	-.002
AA 2 (r)	Benefit Others/Benefit Personally	**-.444**	**.613**	.170
CN 2 (r)	Be defiant/Do my duty	-.247	-.123	**.824**
CN 3	Do what's expected of me/Do the opposite of what's expected of me	**.535**	.177	**.582**
CN 1	Do what I'm supposed to do/Do what I'm not supposed to do	**.546**	.293	**.455**

Bold = significant loadings, TP = Telic/Paratelic, CN = Conformist/Negativist, AA = Autic/Alloic, (r) = Reverse coded item (items were randomized in both the order each anchor was coded and the order each item was presented).

Table 3
Component Reliabilities for the Uncrossed pairs of the Branched RTSM – Initial Three-Factor Solution.

Number of items	Component/ Metamotivational Domain	Cronbach's Alpha
3	Telic-Paratelic	.714
3	Conformist-Negativist	.433
3	Autic-Alloic	.506

vestigate whether a four-component structure was more realistic. For the criteria based on eigenvalues, three factors emerged with values above 1, accounting for 62.0 % of total variance across the nine variables. This amount of variance is less than the levels of >80% (Floyd & Widaman, 1995) or >70% (Jolliffe 2002) of variance that some authors have suggested needs to be accounted for. However, it may not always be realistic to account for such high levels of variance, with Streiner (2003) indicating any level of variance over 50% to be sufficient.

Whether the variance accounted for by non-retained components can be considered less than significant is of equal importance to establishing the acceptable levels of variance explained by a component structure. Rummel (1970) suggested that the retention of components should only cease once the level of "trivial variance" (Ford, MacCallum & Tait, 1986, p. 294) is left unaccounted for. However, interpreting what is and is not "trivial" can prove both ambiguous and highly subjective.

Whilst the variance accounted for by the first two components is clearly significant (32.4% and 15.9% respectively), judging the significance of variance explained by the next

two components is more complex. In the current study, the third and forth components accounted for 13.7% and 9.4% of total variance respectively. It is questionable whether the drop in total variance explained between the two components is sufficient to satisfy Rummel's (1970) protocol for ceasing the retention of factors.

When looking at the proximity to an eigenvalue of 1.0 of the third and forth components in the current study, it could be argued that sound justification is also lacking for only retaining three components. The eigenvalues of the third (principal) and fourth (non-principal) components were 1.23 and 0.85 respectively. Inflexible adherence to the >1.0 eigenvalue rule can be seen as problematic when values are either side of 1.0 and judgments are often arbitrary in nature (Ford et al. 1986). With this in mind, Jolliffe (1972) suggested lowering the eigenvalue level for retaining components to 0.7. So even if a component has an eigenvalue of <1.0 it may still be wise to retain it if can still be said to account for a "significant" amount of variance.

Difficulties in interpreting factors based on the criteria of eigenvalues and variance explained may be overcome through the administration of the scree test (Cattell, 1966). When the scree test was performed on the distribution of factors in the current study, the existence of a four-, rather than three-component solution appeared more prominent. This suggested component solution supports Floyd and Widaman's (1995) assertion that results of a scree test may involve the retention of factors with eigenvalues notably under the value of 1.0.

Although running counter to the theoretical structure assumed by the RTSM, the next stage in the PCA process was to investigate the feasibility of a four-component solution by examining individual factor loadings. In addition to providing the pattern of loadings, examination of component load-

Table 4

Rotated Factor Loadings for the Uncrossed pairs of the Branched RTSM – Forced Four-Factor Solution.

Principal Components

Pair	Item Anchor	Factor Loadings 1	2	3	4
TP 2	Do something serious/Do something playful	**.829**	.142	.006	-.043
TP 3	Accomplish something for the future/Enjoy myself at this moment	**.777**	.021	.248	-.075
TP 1	Do something crucial/Do something of no great concern	**.706**	.139	.223	-.063
AA 1	Focus on my own needs/Focus on others' needs	.179	**.797**	.107	.199
AA 3	Do something for myself/Do something for others	.082	**.771**	.303	-.037
CN 3	Do what's expected of me/Do the opposite of what's expected of me	.205	.133	**.817**	-.027
CN 1	Do what I'm supposed to do/Do what I'm not supposed to do	.247	.269	**.718**	-.031
AA 2 (r)	Benefit Others/Benefit Personally	-.087	.322	-.203	**.836**
CN 2 (r)	Be defiant/Do my duty	-.160	**-.451**	**.441**	**.641**

Table 5

Component Reliabilities for the Uncrossed pairs of the Branched RTSM – Forced Four-Factor Solution.

Number of items	Component/ Metamotivational Domain	Cronbach's Alpha
3	Telic-Paratelic	.714
2	Conformist-Negativist	.621
2	Autic-Alloic	.656
2	Reversed Items	.332

ings can also establish the sufficient worth of items within each component. Item loadings less than 0.40 are generally accepted to be less than significant (Clark & Watson,1995; Floyd & Widaman, 1995; Ford et al. 1988) and so an item must load significantly on one component. Items that cross-load are seen to be lacking in clear distinction of what it is they are measuring, or are indicative of a general factor.

In order to obtain accurate component loadings for a four-component solution, the PCA had to be re-run for the nine variables comprising the uncrossed RTSM branched items (see Tables 4 and 5). This time an extraction based on four components was specified instead of the extraction criteria based on eigenvalues used previously. When examined, the pattern of loadings supported the intended structure of the RTSM for the first three components. The first seven items accounted for in these three components all exhibited loadings >0.70 and no cross-loadings based on criteria of >0.40 for significant loadings.

However, the forth component of the analysis consisting of the second items from the autic-alloic and conformist-negativist pairs respectively, proved a greater challenge to explain. Descriptions of all items included in the analysis were then examined and it was noticed that both items loading onto the forth component (thus separately from the intended theoretical structure) had been reverse coded. That is

to say, the order in which the two sides of the metamotivational pair expressed were presented was opposite to the rest of the items.

Further consultation of psychological and methodological literature reveals the potentially harmful impact that such reverse-coded items can have. It has been suggested that the use of reverse-coded items can have the effect of reducing validity (Schriesheim & Hill, 1981) and producing systematic error in respondents (Jackson, Wall, Martin, & Davids, 1993). Traditionally, reverse-scored items have been employed to address the issue of response pattern bias, based on recommendations from authors such as Likert (1932). However, Hughes (2009) states there to be little empirical support for following this practice. Most tellingly for the current study, both Schmitt and Stults (1985) and Hughes (2009) warn that reverse-coded items will often load on a separate factor or component to all other items in a measure.

With the results of the current study appearing to support this view, the conclusion may be elimination of the two reverse-coded items from the measure. However, in deciding on the most appropriate component structure to adopt, it is important to explore a range of alternatives and to "examine the interpretability of alternative factor solutions" (Floyd & Widaman, 1995, p. 292). It must also be stressed that the result of eliminating any of the nine items under analysis would result in at least one component containing only two variables. Whilst it is possible to justify such a course of action, the resultant creation of a two-item component is less than desirable. Most authors are in agreement that a single component or factor should contain a minimum of three variables (e.g. Cook, Hepworth, Wall, & Warr, 1981; Tabachnick & Fidell, 2013).

With these considerations in mind, the factor loadings of the original three-component solution (based on eigenvalues >1.0) were examined (see Table 2). The (varimax) rotated component solution generally supported the three separate structures of the telic-paratelic, conformist-negativist and autic-alloic pairs of items. However, factor loadings

Table 6

Rotated Factor Loadings for the Uncrossed pairs of the Branched RTSM – Forced Three-Factor Solution with Reverse-Coded Items Removed.

		Principal Components		
		Factor Loadings		
Pair	Item Anchor	1	2	3
TP 2	Do something serious/Do something playful	**.843**	-.027	.212
TP 3	Accomplish something for the future/Enjoy myself at this moment	**.767**	.304	.000
TP 1	Do something crucial/Do something of no great concern	**.694**	.284	.119
CN 3	Do what's expected of me/Do the opposite of what's expected of me	.174	**.849**	.110
CN 1	Do what I'm supposed to do/Do what I'm not supposed to do	.208	**.765**	.238
AA 1	Focus on my own needs/Focus on others' needs	.142	.073	**.869**
AA 3	Do something for myself/Do something for others	.098	.259	**.793**

Table 7

Component Reliabilities for the Uncrossed pairs of the Branched RTSM – Forced Three-Factor Solution with Reverse-Coded Items Removed.

Number of items	Component/ Metamotivational Domain	Cronbach's Alpha
3	Telic-Paratelic	.714
2	Conformist-Negativist	.621
2	Autic-Alloic	.656

were not as defined as clearly as would be preferable, with three items displaying significant loadings on more than one component. Whilst two of these items still exhibited higher loadings on their expected component (autic-alloic 2 and conformist-negativist 3), one item (conformist-negativist 1) did not, loading higher on the telic-paratelic component. It has been suggested that items that load significantly and at approximate levels on more than one component, may need to be eliminated from a measure entirely (Floyd & Widaman, 1995). This brings into question the utility of the three-item measure with the observed pattern of loadings.

Component Reliability

Reliability refers to how well or appropriately related is a set of items within a component or subscale (Cortina, 1993) and how consistent they are in measuring an underlying construct (Henson, 2001). Each of the three components identified through PCA were tested for internal reliability. Results of the analysis were found to be satisfactory for the first component (three telic-paratelic items; $\alpha = .714$) but not for the second and third components (see Table 3). Results for the second (three autic-alloic items; $\alpha = .506$) and third (three conformist-negativist items; $\alpha = .433$) components were indicative of poor reliability (George & Mallery, 2003). However, for both subscales, results showed that removal of one item would increase the reliability of the component (see Ta-

bles 6 and 7). Thus when the second conformist-negativist item was removed the third component's reliability reached $\alpha = .621$ and when the second autic-alloic item was removed, reliability of the component was raised to $\alpha = .656$.

This increase in reliability brings the two components closer to the commonly accepted reliability level of $\alpha = 0.7$ proposed by many authors (e.g. Dekovic, Janssens, & Gerris, 1991; Holden, Fekken, & Cotton, 1991; George & Mallery, 2003). Therefore, these results present a strong case for the removal of two of the nine items analyzed. This proposal is supported by the findings earlier presenting the four-factor solution in which the same two reverse-coded items loaded on a separate component.

Principal Components and Reliability Analysis – Summary

The analysis of principal components within the uncrossed items of the branched RTSM has been conducted with close reference to a number of recommendations presented in the extant literature. In particular, the analysis has sought to draw on four core recommendations made by Ford et al. (1986): Use of the eigenvalue rule, performance of the scree test, the testing of alternative factor solutions, and the use of a pre-existing theoretical model to guide interpretation. The end result of this process has been the retention of the RTSM's existing three-component solution for the uncrossed pairs, but with the removal of the two reverse-coded items. Despite its drawbacks, this nevertheless represents the most satisfactory solution, with two-item components still considered feasible (Raubenheimer, 2004; Worthington &Whittaker, 2006).

Agreeability

Agreeability in the context of this study refers to the level of replication in which the branched and bundled versions of the RTSM classify respondents into the same metamotivational state. Although both versions adopt the same descriptions of each metamotivational state, differences in the way

Table 8

Analysis of Agreement between Branched and Bundled RTSM.

	Metamotivational Domain		
	Telic/ Paratelic	Conformist/ Negativist	Autic-Alloic/ Mastery-Sympathy
No of Participants	306	306	219
% Of Agreement (including that by chance)	81.7%	86.9%	67.6%
% Of Agreement (excluding that by chance)	42.1%	50.7%	16.2%
Kappa Coefficient (K)	.421	.507	.162
Strength of K	Moderate[1]/fair[2]	Moderate[1]/fair[2]	Slight

[1] Landis, J. R., and Koch, G. G. (1977).

[2] Cicchetti, D. V., and Sparrow, S. A. (1981).

the two measures are processed by the respondent and calculated by the researcher mean that identical results should not be assumed. A snapshot of these criteria can be gleaned from the percentage rates of participants classified into the same metamotivational state across each of the three domains for the two versions of the RTSM (see Table 8).

For the telic-paratelic pair, overall classification post-intervention was 82.4% telic and 17.6% paratelic in the branched version and 78.4% telic, 21.6% paratelic in the bundled version. For the conformist-negativist pair it was 82.4% conformist and 17.6% negativist according to the branched RTSM whilst the bundled RTSM produced measurements of 86.3% and 13.7% respectively. Within the crossed transactional pairs the branched version scored as follows; 88.1% autic-mastery, 3.2% alloic-mastery, 3.2% autic-sympathy, 5.5% alloic-sympathy. The bundled version recorded more variation with 68% of participants in the autic-mastery state, 13.7% in alloic-mastery, 2.7% in autic-sympathy and 15.5% in alloic-sympathy.

Although these figures in general suggest the two versions of the RTSM are able to produce similar results, it does not tell us how much of this agreeability may be down to chance. Thus in order for this to be established a test of agreeability using the kappa statistic for categorical data was performed. For the telic-paratelic pair of states 250 out of 306 participants were classified identically (81.7%) in both RTSM versions ($k = .421$, $p < .001$). For the conformist-negativist pair, 266 out of 306 participants (86.9%) were classified into the same states in both RTSM versions ($k = .507$, $p < .001$). Finally for the crossed transactional states, 148 out of 219 respondents were classified identically (67.6%) in both RTSM versions ($k = .162$, $p < .001$).

There is some degree of variation in the interpretation of the kappa coefficient, but the basic premise being the closer the value is to 1.0, the higher the agreeability (Osborne & Costello 2004). In terms of the current study, the result of agreeability analysis on the telic-paratelic and conformist-negativist states can be said to be either moderate (Landis & Koch 1977) or fair (Cicchetti & Sparrow 1981). Less confidence however, can be placed on the agreeability of the branched and bundled RTSM versions regarding the crossed transactional states. For this set of four states the level of agreeability can only be regarded as slight (Landis & Koch 1977), meaning that more synchronicity is perhaps required in this regard.

Discussion

Recommendations Based on Principal Component and Reliability Analysis

The analysis performed on the RTSM in the current study has identified some issues worthy of consideration in possible refinement of the measure. The removal of two items from the measure can be seen as the most significant outcome of the analysis. However, it does not follow that recommendation should be made to avoid use of these items in all subsequent studies. This is because firstly as noted in the literature, it must be recognized that findings from any principal component and reliability analyses are to a degree, specific to the sample subjected to the measure (Wilkinson & Task Force, 1999; Henson, 2001; Streiner, 2003). Secondly and of more acute significance to the RTSM, the poor results attached to the two deleted items may be largely attributable to the reverse coding of the items rather than their content.

It is desirable that all items on the RTSM and in particular the two deleted items in the current study are subject to further psychometric testing. However, full use of all items on the RTSM is encouraged. Rather, what researchers administering the RTSM should be recommended against employing, is the use of reverse-coded items on any pen and paper administrations of the RTSM. Since the current study did not make use of any on-line administrations of the RTSM it is not clear if the issues with reverse-coded items apply in computer-based contexts. Examination of the Desselles et al. (2014) study, which administered 54.9% of surveys online, shows no reporting of any problems from the use of reverse-coded items.

It must also be recognised that the flexibility of approaches the RTSM allows the researcher also involves the production of categorical data, which cannot be subjected to the previously mentioned analyses. For researchers administering the bundled version of the RTSM though, the question of construct validity still remains crucial to the generation of meaningful data. To this end, a further recommendation is that future testing of the RTSM uses both the branched and bundled versions. If the construct validity and internal consistency of the branched RTSM can be established, then agreeability analysis between the branched and bundled versions can serve to provide a form of cross-validation.

Principal Value of the RTSM

Despite the problems encountered using both versions of the RTSM, the instrument was nevertheless able to effectively capture one of the fundamental requirements of reversal theory research; the phenomenon of reversals. Employing the bundled RTSM as a repeated measure allowed the study to record a significant proportion of metamotivational reversals across the sample, lending empirical weight to reversal theory's central proposition.

Use of the RTSM in this way can be seen as an important advancement of the work by Desselles et al. (2014). Of course it must be recognized that the aim of the authors was not to track reversals but to find accurate measures of metamotivational state. However, empirical investigation directed at the heart of reversal theory requires repeated documenting of metamotivational states. Thus this study has sought to help answer Apter's (2013) call for more research that is able to empirically test the theory.

In addition to allowing the testing of reversals to be made, the RTSM provides a conduit for more comprehensive measuring of metamotivational states. Up until now, state measures within reversal theory have been dominated by the telic-paratelic pair (Svebak & Murgatroyd, 1985; Calhoun, 1995; O'Connell & Calhoun, 2001), the conformist-negativistic pair (O'Connor, 1992) and an overall focus on the somatic states (Cook, Gerkovich, Potocky, & O'Connell, 1993). Thus the RTSM provides the theory with an opportunity to develop a body of research that can enhance understanding around the full range of states. This would help to correct the historical bias in reversal theory research on the telic-paratelic pair. In particular it opens up greater scope to research the transactional pairs and produce more "theory-centred research" (Apter, 2013, p. 1).

The RTSM represents a major breakthrough in reversal theory research that builds on previous measures that did not account for all eight metamotivational domains. The fact that the RTSM is a state rather than trait measure makes it highly sensitive to the empirical needs of the theory; that is to say the moment to moment measuring of states and reversals. However, the RTSM also provides the empirical flexibility to collate changes in metamotivational state over time and produce what Apter (2013) has called "state balance measures" (p. 4). Employing the RTSM in this way would allow research to better track state dominance over time based on phenomenological experience rather than trait inclinations.

Limitations

A notable limitation of the current research concerns the completion rates of questionnaires provided by participants in the study. Both versions of the RTSM contained completed (post-intervention) responses to all items from only just over half of the participants questioned (56.6% for the branched, 54.7% for the bundled). These figures are similar to average survey response rates in the behavioral sciences found in studies by Baruch (1999) and Baruch and Holtom (2008). However, in both studies the authors state that these figures deliberately omitted surveys where questionnaires were administered by the researcher directly to the sample. For surveys administered in this way (such as the current study) the authors state that completion rates are likely to be far higher due to the increased compulsion for participants to take part.

Unlike the majority of studies examined by Baruch (1999) and Baruch and Holtom (2008), the percentage of unusable questionnaires in the current study was not the result of participants failing to return them. Instead, the current study received a number of incorrectly or incompletely returned questionnaires. In particular, the crossed transactional items on the bundled version of the RTSM appeared to pose a problem for participants and recorded the lowest number of correct responses for individual items.

Whether this effect is peculiar to the sample used in the study or resultant from the pen and paper method of administration of the measure is uncertain. Use of the RTSM in a dynamically programmed online format may have improved the completion rate of the current study. However, when administered online, the level of control and variability of the conditions in which questions are answered is dramatically altered. This is particularly sensitive to the framework of reversal theory since the conditions in which a survey is completed are likely to have an effect on the concurrent state being experienced.

Future Research

Two key issues encountered during the completion of the current study need to be put into context within wider use of the RTSM. Firstly, the low completion rates for the crossed pairs in the bundled version and the branched version as a whole need to be further assessed. Secondly, impact of reverse-coded items also needs to be further considered.

Future studies employing the RTSM should seek to administer the RTSM in both dynamically programmed and pen and paper formats. This would allow the merits of each

format to be further explored and differences in completion rates and "pre-score" states to be compared. It needs to be established whether both the issues previously addressed are peculiar to pen and paper administering of the RTSM. If this is so, recommendations about how the RTSM is administered need to be clarified.

Future research should also seek to widen the demographic samples on which the RTSM is used. The current study has developed this with respect to ethnicity. However, one notable variable for future studies to develop is that of age since this and the Desselles et al study (2014) both employed undergraduate samples. There may still be some teething problems with use of the RTSM as evidenced in this study. However, whilst these need to be assessed with further research, it is the capturing of state data across a variety of contexts and samples that should be most embraced by researchers in reversal theory and beyond.

References

Apter, M. J. (2013). Developing Reversal Theory: Some suggestions for future Research. *Journal of Motivation, Emotion, and Personality, 1*, 1-8. doi: 10.12689/jmep.2013.101

Apter, M. J., Mallows, R. & Williams, S. (1998). The development of the Motivational Style Profile. Personality and Individual Differences, 24, 7-18.

Baruch, Y. (1999). Response Rates In Academic Studies – A Comparative Analysis. *Human Relations*, *52*, 421-438. doi:10.1177/001872679905200401

Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, *61*, 1139-1160. doi: 10.1177/0018726708094863

Cattell, R. B. (1966). The Scree Test for the Number of Factors. *Multivariate Behavioral Research*, *1*, 245-276. doi:10.1207/s15327906mbr0102_10

Calhoun, J. E. (1996). *Construct validity of the Telic/Paratelic State Instrument: a measure of reversal theory constructs* (Unpublished Doctoral dissertation). University of Kansas, Lawrence.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing inter-rater reliability of specific items: applications to assessment of adaptive behavior. *American journal of mental deficiency*, *86*, 127-137.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*, 309-319. doi:10.1037/1040-3590.7.3.309

Cliff, N. (1988). The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin*, *103*, 276-279. doi:10.1037/0033-2909.103.2.276

Comfrey, A. L., & Lee, H. B. (1992). *A First Course in Factor Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, J.D., Hepworth, S.J., Wall, T.D. & Warr, P.B. (1981). *The Experience of Work: A Compendium and Review of 249 Measures and their Use*. San Diego: Academic Press.

Cook, M. R., Gerkovich, M. M., Potocky, M., & O'Connell, K. A. (1993). Instruments for the assessment of reversal theory states. *Patient Education and Counseling*, *22*, 99-106. doi:10.1016/0738-3991(93)90006-I

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 98-104. doi:10.1037/0021-9010.78.1.98

Dekovic, M., Janssens, J. M., & Gerris, J. R. (1991). Factor structure and construct validity of the Block Child Rearing Practices Report (CRPR). *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*, 182-187. doi:10.1037/1040-3590.3.2.182

Desselles, M. L., & Murphy, S. L. (2011). *Development of a Comprehensive State Measure*. Paper presented at the Fifteenth International Reversal Theory Conference, Alexandria, VA. Abstract retrieved from http://www.reversaltheory.net/files/2011/Reversal-Theory-Conf-Book.pdf

Desselles, M. L., Murphy, S. L., & Theys, E. R. (2014). The Development of the Reversal Theory State Measure. *Journal of Motivation, Emotion, and Personality, 2*, 10-21. doi:10.12689/jmep.2014.202

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272. doi:10.1037/1082-989X.4.3.272

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*, 286-299. doi:10.1037/1040-3590.7.3.286

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The Application of Exploratory Factor Analysis in Applied Psychology: A Critical Review and Analysis. *Personnel Psychology*, *39*, 291-314. doi:10.1111/j.1744-6570.1986.tb00583.x

George, D., & Mallery, M. (2003). Using SPSS for Windows Step by Step: A Simple Guide and Reference. Boston, MA: Allyn & Bacon.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, *34*, 177-89.

Holden, R. R., Fekken, G. C., & Cotton, D. H. (1991). Assessing psychopathology using structured Test-item response latencies. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*, 111-118. doi: 10.1037/1040-3590.3.1.111

Hughes, G. D. (2009). The Impact of Incorrect Responses to Reverse-Coded Survey Items. *Research in the Schools*, *16*, 76-88.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*, 401- 415. doi:10.1007/BF02291817

Jackson, P. R., Wall, T. D., Martin, R., & Davids, K. (1993). New measures of job control, cognitive demand, and production responsibility. *Journal of Applied Psychology*, *78*, 753. doi:10.1037/0021-9010.78.5.753

Jolliffe, I.T. (1972). Discarding Variables in a Principal Component Analysis 1: Artificial Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *21*, 160–173.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd Edition). New York: Springer.

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, *33*, 159-174.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 4-55.

McDermott, M. R. (1988). Measuring rebelliousness: The development of the Negativism Dominance Scale, in: Apter, M. J., Kerr, J. H. & Cowles, M. P. (Eds). Progress in Reversal Theory, Elsevier: Amsterdam.

Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw-Hill.

O'Connell, K. A., & Brooks, E. (1997). Resisting urges and adopting new behaviors. In S. Svebak & M. J. Apter (Eds.), Stress and health: A reversal theory perspective (pp. 157-171). Washington, DC: Taylor & Francis.

O'Connell, K. A., & Calhoun, J. E. (2001). The telic/paratelic state instrument (T/PSI): validating a reversal theory measure. *Personality and Individual Differences, 30,* 193-204. doi:10.1016/S0191-8869(00)00026-X

O'Connor, P. A. (1992). *Reversal theory and mother–child compatibility.* (Unpublished doctoral dissertation). University of Tasmania, Australia.

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation*, 9.). Retrieved from http://pareonline.net/Home.htm

Peterson, R. A. (1994). A meta-analysis of Cronbach's Coefficient Alpha. *Journal of Consumer Research*, *21*, 381-391.

Raubenheimer, J. (2004). An item selection procedure to maximize scale reliability and validity. *SA Journal of Industrial Psychology*, *30*, 59-64. Retrieved from http://reference.sabinet.co.za/sa_epublication/psyc

Rummel, R. J. (1970). *Applied Factor Analysis*. Evanston, IL: Northwestern University Press.

Schmitt, N., & Stults, D. M. (1985). Factors Defined by Negatively Keyed Items: The Result of Careless Respondents? *Applied Psychological Measurement*, *9*, 367-373. doi: 10.1177/014662168500900405

Schriesheim, C. A., & Hill, K. D. (1981). Controlling Acquiescence Response Bias by Item Reversals: The Effect on Questionnaire Validity. *Educational and Psychological Measurement*, *41*, 1101-1114. doi:10.1177/001316448104100420

Streiner, D. L. (2003). Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, *80*, 99- 103. doi:10.1207/S15327752JPA8001_18

Svebak, S., & Murgatroyd, S. (1985). Metamotivational dominance: A multimethod validation of reversal theory constructs. *Journal of Personality and Social Psychology*, *48*, 107. doi: 10.1037/0022-3514.48.1.107

Tabachnick, B. G., & Fidell, L. S. (2013). Using Multivariate Statistics (6th Edition). New York: Harper & Row.

Tucker, J. (2012). *Reversal Theory Training Site - Sample Videos.* Available from http://reversaltheory.net/RT/reversal-theory- training

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology, 54,* 1063-1070. doi:10.1037/0022-3514.54.6.1063 Wilkinson, L. & the Task Force on Statistical Inference, American Psychological Association, Science Directorate. (1999). Statistical Methods in Psychology Journals: Guidelines and Explanations. *American Psychologist*, *54*, 594- 604. doi: 10.1037/0003-066X.54.8.594

Worthington, R. L., & Whittaker, T. A. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist, 34*, 806-838. doi:10.1177/0011000006288127

Young, T., Desselles, M., Lee, J., & Apter, M. J. (2005, July). *Measuring reversal theory states in real time: Developing the Reversal Theory State Inventory.* Paper presented at the 12th International Reversal Theory Conference, Winnipeg, Canada. Abstract retrieved from http://www.reversaltheory.net/files/2005-abstracts.pdf