Motivation and Emotion
https://doi.org/10.1007/s11031-020-09831-9

ORIGINAL PAPER

# A review and empirical comparison of motivation scoring methods: An application to self-determination theory

Joshua L. Howard[1] · Marylène Gagné[2,3] · Anja Van den Broeck[4,5] · Frédéric Guay[6] · Nikos Chatzisarantis[7] · Nikos Ntoumanis[7] · Luc G. Pelletier[8]



## Abstract

Self-determination Theory differentiates various types of motivation, each of which have different consequences for well-being and behavior. Despite broad agreement concerning the nature of different types of motivation, numerous scoring methods, each of which rely on different assumptions, are commonly practiced. These practices range from a relative autonomy index that collapses all types of motivation into a single index, higher-order models grouping subscales into a two-factor solution, to multi-factorial approaches examining all motivation types as separate constructs. Existing evidence has not empirically compared these methods or clearly favored the use of one over another. We review each method and further investigate the advantages and disadvantages of each approach by directly comparing a range of commonly utilized scoring methods, as well as recently developed methods across six independent samples from various life domains to determine their effectiveness. Results generally favor multidimensional methods (e.g., exploratory structural equation modeling, B-ESEM, and CFA) as more comprehensive scoring practices as they maximize construct relevant information. However, selection of an ideal method will rely on theoretical congruence between methodology and research questions.

**Keywords** Self-determination theory · Motivation · Higher-order · RAI · Bifactor · ESEM

## Introduction

Self-determination theory (SDT; Ryan and Deci 2017) is widely used to understand human motivation. It is based upon the proposition that motivation is not a unidimensional construct that varies in terms of how much motivation people possess, but rather is multidimensional and best represented by distinct types of motivation (or regulations) ranging from not being motivated at all (amotivation), being motivated by external incentives, to intrinsic motivation. SDT's types of motivation have been studied across various contexts ranging from education, sport, exercise, health, volunteering and work, to more specific areas including knowledge sharing, computer gaming, friendship, parenting, and political engagement. This body of research has produced knowledge about the importance of differing types of motivation for behavioral engagement and well-being (Ryan and Deci 2017).

Due to the complex theoretical conceptualization of motivation and associated difficulties with hypothesis testing often involving up to six separate motivational constructs, different methods for aggregating motivation types have

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s11031-020-09831-9) contains supplementary material, which is available to authorized users.

✉ Joshua L. Howard
josh.howard@monash.edu

1 Monash Business School, Monash University, Melbourne, VIC, Australia

2 Future of Work Institute, Curtin Business School, Curtin University, Perth, Australia

3 University of Western Australia, Perth, Australia

4 Department of Work and Organization Studies, KU Leuven, Leuven, Belgium

5 Optentia, North-West University, Vanderbijlpark, South Africa

6 Department of Educational Fundamentals and Practices, Université Laval, Québec, Canada

7 School of Psychology, Curtin University, Perth, Australia

8 Human Motivation Research Laboratory, School of Psychology, University of Ottawa, Ottawa, Canada

Published online: 17 April 2020

been developed. However, little guidance exists to suggest when each method is preferable, and even less empirical work has been conducted to compare the adequacy and implications of applying one method relative to another. This is notable because improving the conceptualization of psychological constructs (Reeve 2016) and ensuring validly of measurements are essential, yet often times overlooked requirements of psychological research. As a result, while each method has demonstrated utility in various areas, the implications of applying one method over another remain unclear. The use of these different scoring methods has a potential impact on research conclusions as each computational method relies on different theoretical assumptions. Likewise, some methods may be statistically or practically better supported than others, and therefore predict different types of outcomes more effectively. In this paper, we begin by reviewing common scoring practices and examine their theoretical support before empirically comparing these methods through primary and archival datasets from multiple countries, languages, and life domains. Specifically, we focus on which methods perform better in predicting outcomes while taking into account statistical issues. We then discuss the suitability of each method and offer recommendations concerning how to optimally measure and model motivation types.

This study has both theoretical and practical merits for those studying SDT. From a theoretical point of view, we highlight the underlying assumptions and theoretical importance of various scoring methods which will enable motivation to be operationalized in optimal ways. From a practical perspective we also offer recommendations and solutions to common problems associated with scoring methods (e.g. multicollinearity) which can be applied to studies both within SDT and other areas of research. Although not the focus of this study, our results apply equally to other multidimensional constructs and as such may inform the use of different scoring methods for constructs in other theoretical areas.

## Self-determination theory

Self-determination theory (Ryan and Deci 2017) specifies a range of different behavioral regulations, each of which captures different reasons why people might choose to pursue activities. These regulations are theoretically (Ryan and Deci 2017) and empirically distinct (e.g. Gagné et al. 2015; Vallerand et al. 1992; Pelletier et al. 2013), yet can also be predictably ordered along a continuum of self-determination (Ryan and Deci 2017; Chemolli and Gagné 2014). Specifically, regulations are consistently ordered from most self-determined (intrinsic motivation) to least self-determined (amotivation; see Howard et al. 2017; Sheldon et al. 2017).

The most self-determined, intrinsic motivation is evident when actions are pursued solely due to the enjoyment and interest inherent in engaging in the behavior. Second, integrated regulation describes cases in which the enactment of a behavior is assimilated into the individuals' sense of self, and thereby becomes a part of their ongoing identity. While theorized and occasionally measured, this form of motivation is often excluded during scale development due to a lack of discriminant validity (Gagné et al. 2015; Howard et al. 2017). Identified regulation refers to behavior pursued for the meaning an individual experiences while engaging in a behavior. For example, while being environmentally responsible may not be inherently enjoyable (intrinsic motivation), people pursue environmentally friendly initiatives because they perceive them as a highly meaningful (identified regulation; Pelletier et al. 1998).

Whereas intrinsic, integrated, and identified regulations are considered autonomous, the next two regulations are considered controlled forms of motivation. Introjected regulation refers to the pursuit of a behavior in order to attain or avoid personally-administered rewards or punishments in the form of ego-involved emotions, such as pride, shame, and guilt. For example, an employee may feel guilty after underperforming at work and stay at the office after standard work hours. External regulation describes behavior which is encouraged through rewards (e.g. monetary incentives) or punishments administered by others, for example, when an employee works hard to reach a quota that leads to a bonus. Finally, amotivation is described as a lack of intention to enact a behavior (Ryan and Deci 2000). There is some contention about whether amotivation should be considered within SDT given that it is, by definition, a lack of intentional motivation, or lack of understanding behind one's reason for acting (Chatzisarantis et al. 2003). However, amotivation is often included in SDT scales (e.g. Gagné et al. 2015), appears to fit the SDT structure as demonstrated through multidimensional scaling (Howard et al. 2017; Sheldon et al. 2017), and often explains a non-negligible portion of variance in outcomes. It is therefore considered in this study.

This complex representation of motivation in which regulation types are categorically distinct and yet also interdependent as part of a single continuum has no doubt contributed to the proliferation of scoring methods. Each scoring method can be justified by parts of the theory, yet each method relies on different theoretical assumptions. The practice of collapsing regulations into a composite factor (e.g. a relative autonomy index; Grolnick and Ryan 1989; Sheldon et al. 2017), as a telling example, is justified by the continuum logic in which the shared elements of each regulation are emphasized, and the distinguishing characteristics of individual regulations as separate categories are de-emphasized. Alternatively, approaches that use each regulation as a distinct variable specifically emphasize the unique characteristics of each motivation type, yet largely ignore their shared variance. Though both approaches are widely used (among others), they are not operationalizing motivation in the same way, with each having their own advantages and disadvantages. Yet, a thorough comparison of the implications of applying one method over another is currently lacking. In the following section we describe the different scoring methods that have been used in SDT and discuss their advantages and disadvantages.

## Scoring methods

*The Relative Autonomy Index* (RAI; Grolnick and Ryan 1989) is a scoring method in which each regulation subscale is calculated before being weighted and combined with other regulations according to their assumed position on the SDT-continuum. The result is a single score representing the degree of relative autonomy. While the formulation of this score changes depending on the regulations measured, the general formula is as follows:

$$RAI = (-2 * External) + (-1 * Introjected) + (1 * Identified) + (2 * Intrinsic)$$

The RAI has been one of, if not the most popular scoring method due to its practicality and has contributed to demonstrating the value of self-determination theory's proposition that the relative level of autonomous motivation is essential in predicting outcomes (Ryan and Deci 2017). The RAI is justified based upon the continuum logic in which each type of motivation is predictably ordered and can be classified as either positive or negative, as denoted by the weightings, with external and introjected regulations reducing the degree of autonomy and identified and intrinsic motives increasing the level of relative autonomy. By adding these positively and negatively weighted scores together, this index estimates the overall degree of relative autonomy and as such approximates an individuals' position along the underlying continuum of self-determination. Given the overwhelming evidence that the degree of self-determination is an important factor in ones' motivation (Howard et al. 2018; Litalien et al. 2017; Ryan and Connell 1989; Ryan and Deci 2017), the RAI has proved popular and useful for testing hypotheses relating to the degree of self-determination.

However, while the RAI theoretically aligns with the continuum of self-determination, it does not align with SDT's premise that motivation is multidimensional and that each type of regulation will relate to covariates differently. By combining all subscales into a single index, the RAI is no longer able to detect what role each individual type of motivation is playing in a given analysis. For example, by examining the full range of motivation types in relation to job proficiency, proactivity, academic performance, and voting behaviors, it was found that identified regulation was as important, if not more important than intrinsic motivation (Burton et al. 2006; Gagné et al. 2015; Losier and Koestner 1999). Moreover, a longitudinal study of competitive swimmers demonstrated that while introjection predicted short-term persistence over a period of one year, it was not significantly related to this outcome variable over a period of two years. In contrast, external regulation did not display significant effects over a period of one year but proved negatively associated with persistence over a period of two years, showing that these types of motivation have different longitudinal effects, and therefore cannot safely be combined without obscuring pertinent findings (Pelletier et al. 2001). The RAI is not able to detect such results and instead assumes that intrinsic motivation will always be more important than identified regulation, and that effects will be consistent over time, regardless of what the covariates may be. Likewise, the RAI assumes that introjected regulation is an entirely negative force, as indicated by its negative weighting, despite evidence indicating it can at times relate positively to desirable outcomes such as positive affect and enactment of exercise behaviors (Ng et al. 2012). These considerations are important as they demonstrate that the RAI is likely to exclude information that may be important, both theoretically and practically (Bono and Judge 2003; Edwards 2001; Howard et al. 2018; Litalien et al. 2017).

Additionally, the RAI also entails considerable statistical limitations. First, the weights associated with each subscale are relatively arbitrary with no published empirical evidence to support them (Ryan and Connell 1989; Grolnick and Ryan 1989). Interestingly, recent evidence has suggested the weights themselves might not matter (Sheldon et al. 2017). Specifically, Sheldon and colleagues tested several alternate calculation methods, including one in which no weighting was used in RAI calculation, and found that the predictive validity was unchanged regardless of weighting system. Secondly, the RAI is a difference score and as such inherits problems associated with these methods, including low reliability (Edwards 2001; Johns 1981). Third, it is worth noting that this process typically relies on averaged scale scores being calculated before combining each subscales. However, alternate methods for calculating similar indices of self-determination exist (e.g. Guay et al. 2003), though all are likely to exclude information associated with regulation types.

Taken together, the RAI approximates a score of an individuals' level of self-determination and as such may be applicable when testing simple hypotheses concerning this continuum factor. However, it is insensitive to unique effects associated with individual regulation types, and as such fails to account for the multidimensional nature of motivation detailed in SDT. Given this, while we expect the RAI will predict outcomes moderately well, it not be able to explain as much variance in outcomes when compared to more complex scoring methods as the construct-relevant information concerning unique effects of individual regulation will not be modeled.

### Higher-order models

In the higher-order approach, two factors are calculated: one combining intrinsic and identified regulations to form a factor of autonomous motivation, and one combining introjected and external regulations to form a factor of controlled motivation. When amotivation is measured, it is generally represented as a separate latent factor (Gagné et al. 2015; Li 1999), whereas integrated regulation, when measured, is combined into the autonomous factor. Within higher-order models, motivation is defined primarily by degree of autonomy with "autonomous motivation" representing a high degree of self-determination, whereas "controlled motivation" primarily models a low or negative degree of self-determination. Like the RAI, this method does not fully account for the unique characteristics of regulation subscales. However, unlike the RAI, it does give a dualistic perspective of self-determination by individually examining the effects of both the presence of self-determination (autonomous motivation), and absence of self-determination (controlled motivation). This allows for a more nuanced and potentially theoretically important examination of relative autonomy and allows for examination of both direct effects and interaction effects (e.g. Phillips and Johnson 2018).

By specifying two factors that are most often uncorrelated, this method appeals to researchers because of its parsimony and reduced issues of multicollinearity, therefore yielding easily interpretable results. For example, a typical application of this method may demonstrate a negative relation between autonomous motivation and turnover intentions, while controlled motivation may relate positively to the same outcome (e.g., Gillet et al. 2013). It is also worth noting that while higher-order models can be specified through several methods, we refer here to the true specification in which items load onto a first-order latent factor representing each subscale, before these latent subscale factors then load onto second-order factors representing autonomous and controlled motivation factors. Other practices, for example creating subscale scores through first averaging items before loading these subscales onto first-order autonomous/controlled factors, do not control for measurement error as rigorously.

However, higher-order models are not well supported on either theoretical or empirical grounds. From a theoretical perspective, if a continuum underlies the regulation types and each is relatively equally spaced along this continuum (as has been demonstrated empirically; Howard et al. 2017; Sheldon et al. 2017), then there is no clear justification for the division between autonomous and controlled factors to be placed between introjected and identified regulations. Meta-analytic and multidimensional scaling studies have demonstrated that introjection is equally correlated with external and identified regulations ($r = 0.60$ in both cases; Howard et al. 2017) questioning why it should be grouped with external regulation. Furthermore, recent research using other statistical methods (i.e., bifactor-ESEM) has not supported the clustering of external and introjected regulations as these variables did not show similar loading onto the general self-determination latent variable (Howard et al. 2018; Litalien et al. 2017), indicating these types of motivation are characterized by substantially different degrees of self-determination.

Alternately, if we consider the theoretical position that motivation types are categorically different, contain unique properties, and will associate with covariates differently, then collapsing these regulations into higher-order factors does not seem theoretically consistent. In fact, the effects of external and introjected regulation are demonstrably different with meta-analytic evidence indicating that, for example, introjected regulation correlates positively with need satisfaction, autonomy support, and positive affect whereas external regulation relates negatively to each of these covariates (Ng et al. 2012; Vasconcellos et al. 2019). As documented above, identified and intrinsic motives have likewise demonstrated differential prediction for certain outcomes (Burton et al. 2006; Losier and Koestner 1999). This again indicates that relevant regulation-specific information will be systematically excluded when constructing high-order models.

Given these considerations, while higher-order models are convenient and can demonstrate the differences between high levels of autonomy compared against low levels of autonomy, this method is not closely aligned with theory or well supported empirical evidence. Like the RAI, this approach is likely to exclude regulation specific information. As such, we predict these models will not capture as much information as more complex methods and instead will perform approximately as well as the RAI.

### Latent subscale factors

The next approach is the calculation of individual motivation subscales and subsequent application of all available subscales to analyses (e.g., correlation, multiple regression). While traditionally these subscales may have been calculated using an average of their items, research has moved to more sophisticated analyses (i.e., Confirmatory Factor Analysis; CFA), which include benefits such as accounting for measurement error. The advantages of the subscale approach lie in its comprehensiveness. By using all available motivation subscales, CFA solutions allows all possible construct-relevant information to be modeled. Unlike the other methods described above, the subscale approach makes full use of the multidimensional conceptualization of motivation. Specifically, when using subscales, motivation is operationalized as a multidimensional construct which assumes each subscale will contain unique characteristics, which in turn will result in different effects when associated with covariates. For this reason, the subscale approach is likely to increase the overall predictive strength of models as well as capture more nuanced results, as demonstrated by a meta-analysis of SDT in the health domain (Ng et al. 2012). For example, whereas introjection is typically classified as a controlling type of motivation and therefore expected to produce negative effects, results have demonstrated that in addition to expected results (e.g., positive relationships with burnout and anxiety), introjection also relates positively to engagement in health behaviors (Ng et al. 2012), job satisfaction, and effort (Gagné et al. 2015). In addition to the above-noted differences between identified and intrinsic motivations, such results further highlight subscale-specific findings that would not be detected by more simplified measures.

While the subscale approach derives benefits from its greater emphasis on the multidimensional nature of motivation, it also downplays the importance of the underlying continuum of self-determination. Specifically, while the subscale approach does model the continuum through correlated factors, it does not distinguish this continuum factor as a distinct variable, and therefore conflates the influence of unique motivation characteristics and the degree of self-determination contained within each regulation. This in turn results in inflated inter-regulation correlations which can cause issues of multicollinearity when applied in subsequent analyses. For example, when predictors are correlated (as they are with SDT-based regulations; Howard et al. 2017), the explained variance can be partitioned incorrectly between predictors due to multicollinearity (Tonidandel and LeBreton 2011), potentially resulting in erroneous effect sizes and even suppression effects in which effects are estimated in the opposite direction to expectation. If not addressed, this can lead to erroneous conclusions concerning the importance of each of the regulations.

Taken together, the subscale approach is more comprehensive than previously reviewed methods and as it will likely capture more construct-relevant information. Therefore, we expect it will account for more variance in outcomes than either RAI or higher-order methods. However, we also expect to see the effects of multicollinearity when examining results of regression analyses.

### ESEM and bifactor models

Finally, two more advanced versions of latent factor models have recently been introduced in SDT research: exploratory structural equation modelling (ESEM) and bifactor-ESEM (B-ESEM). ESEM combines the benefits of traditional CFA and EFA procedures (Asparouhov and Muthén 2009; Marsh et al. 2014; Morin et al. 2013) and allows for cross-loadings to be modeled on each factor of a multidimensional scale. In doing so it maximizes construct relevant information captured by these factors and minimizes bias in the estimation of factor loadings and latent factor inter-correlations. Further, the application of target rotation in ESEM allows for models to be confirmatory in nature by targeting cross-loadings to be as close to zero as possible and penalizing divergence from zero by reducing model fit (Marsh et al. 2014; Morin et al. 2016a, b). Recent research has demonstrated the value of this approach as it better distinguishes between the regulation factors of SDT (i.e. lower inter-factor correlations; Guay et al. 2015; Howard et al. 2018; Litalien et al. 2017). As such, this method is highly similar to the CFA subscales approach detailed above, both theoretically and empirically, except that it will better deal with multicollinearity due to reduced inter-regulation correlations. As such, we predict it will account for approximately the same amount of variance in outcomes as CFA-derived subscales.

B-ESEM is an extension of this ESEM framework to include a general (G-) factor in addition to the regulation specific (S-) factors. This method allows researchers to examine the role of both the degree of self-determination in a person's total motivation (i.e., the G-factor) as well as the unique characteristics of each regulation once the degree of self-determination has been removed (i.e., S-factors). Previous research has identified the G-factor in these models as degree of self-determination, rather than amount of motivation (Howard et al. 2018; Litalien et al. 2017). This assertion is based upon factor loadings which were strongly positive for intrinsic and identified items, moderately positive for introjected items, weakly positive for external regulation items, and moderately negative for amotivation items. Factor loadings would be of more equal value and in the same direction across regulation items if this factor represented general motivation or noise variables such as common method variance. This bifactor method shows potential for significant theoretical progress as this method alone is able to model both the theoretical continuum while simultaneously maintaining the multidimensional nature of motivation specified within SDT literature. As such, this method can fully model the complexity of motivation within SDT. Studies applying this method by Howard et al. (2018) and Litalien et al. (2017) found a G-factor primarily representing the degree of self-determination to be the most important predictor of workplace and educational outcomes, but also found that individual regulations explained some variance in these outcomes beyond the G-factor. It is also worth noting that this method is not influenced by multicollinearity as are the other multidimensional methods reviewed above. Instead, in B-ESEM all commonality between subscales is modeled through the general factor and inter-factor correlations are constrained to zero. While well suited to modeling motivation within SDT, the complexity of B-ESEM models means that samples must be sufficiently large in order to conduct analyses, providing a limitation on this method's applicability. While avoiding the major limitation of other multidimensional methods, this method is complex and requires many factors in order to model motivation, potentially resulting in large and multifaceted modeling when integrated with covariates.

### Overview

This study was designed to review common motivation scoring methods within SDT and to highlight differences between them in terms of predictive validity, while also taking into consideration statistical soundness, parsimony, and theoretical congruity. Specifically, we empirically compared each scoring method in terms of (a) goodness of model fit including consideration of parsimony, and (b) ability to predict outcomes. Goodness of model fit is important because it indicates the degree to which the specified model accurately represents the data. Parsimony indicators highlight whether the addition of more parameters (i.e. greater complexity) is justified by the information the more complex model captures. That is, if models are needlessly complex, fit statistics that account for parsimony will indicate their unsuitability and favor more simple models. Finally, we examine the total variance explained in outcomes as an indicator of how much construct-relevant information is modeled through each scoring method. When two scoring methods are calculated from the same data, yet one predicts substantially more variance in outcomes, this indicates a difference in the degree of construct-relevant information captured by each scoring method. These indicators are tested across six samples collected from different life domains, countries, and with various motivation scales. A graphical representation of measurement models tested, excluding the RAI, is provided in Fig. 1. Combining this information with consideration of the theoretical and conceptual meaning behind each method will allow researchers to make better informed decisions when selecting scoring methods.

## Method

### Participants and procedure

Data consisted of six samples sourced from different domains of research (work, volunteer, education, exercise, physical education, and sport) and from a range of countries

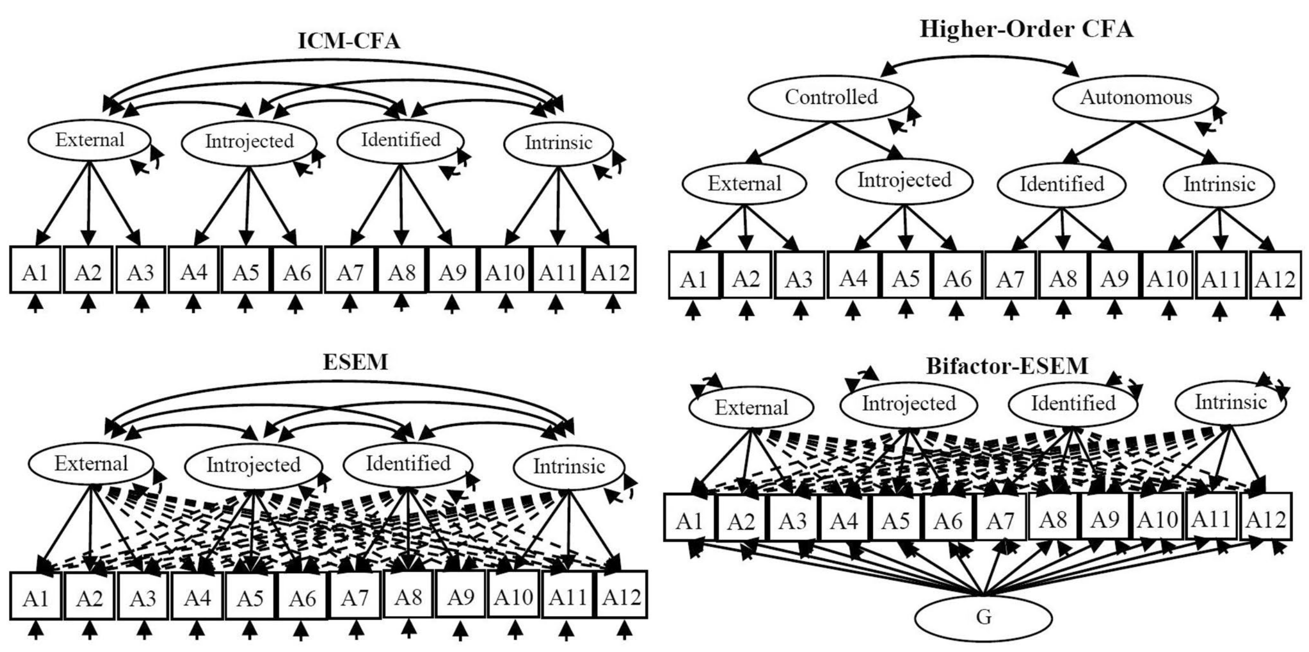


**Fig. 1** Graphical representation of tested models. *CFA* confirmatory factor analyses, *ICM* independent cluster model, *ESEM* exploratory structural equation modeling, *G* global factor, *A*1–*A*12 items; full unidirectional arrows represent main factor loadings; dotted unidirectional arrows represent cross-loadings

(Belgium, the Netherlands, Australia, Canada, USA, and Greece). All samples, with exception of Sample 2 (volunteer) and Sample 4 (exercise), have been previously used in publications (See Table S1 in the online supplementary materials for publication details).

Sample 1 (workplace sample) included 621 participants answering questions concerning their motivation at work. They were recruited from different workplaces in Flanders, Belgium and completed either paper or online surveys in 2009. Participants were on average 37.61 years old (SD = 10.21 years, female = 53.1%).

Sample 2 (volunteer sample) was collected in Australia in 2014 as part of a larger and ongoing study. Australian adult volunteers from a single volunteer-based organization (N = 210) completed an online questionnaire in English. Participants were primarily male (61.6%) and 35.5% of the sample reported being in the age category of 45–54 years old.

Sample 3 (education sample) was collected from high school students in Canada between 2003 and 2005. Surveys, administered in French, were originally completed in paper during class, with second and third time points administered online. At the first time point, 941 student responses (417 male; 524 female) were recorded, with 830 students completing the Time 2 survey, and a further 776 students completing the Time 3 administration. Students were from grade 8 (n = 309), grade 9 (n = 272) and grade 10 (n = 272), with a mean age of 14.83 (SD = 9.61) at time 1.

Sample 4 (exercise sample) consists of 233 university students who completed a paper and pencil survey asking questions about leisure-time physical activities in the United States. The mean age was 18.12 (SD = 4.97) and the sample consisted of 53% males. This study was prospective with psychological variables being measured at three points in time over an 8-week period.

Sample 5 (physical education sample) contains responses from 419 Greek middle school students over three time points from surveys administered in class between November and April of 2014. The mean age of the sample was 14.5 years (SD = 0.53), and 52.6% of participants were male. One, two, and nine students missed the first, second, and third wave of data collection respectively.

Sample 6 (sport sample) consisted of 241 Canadian athletes measured at a single time in 2017, with measures administered online and in French. On average, the participants were 21.49 years old (SD = 5.60 years) and the majority were female (64.3%). The sample was evenly split between athletes competing in individual (46%) and team sports (54%).

### Measures

Sample 1 completed the Multidimensional Work Motivation Scale (MWMS; Gagné et al. 2015), which measures amotivation, external material and external social, in addition to introjected, and identified regulations, and intrinsic motivation. Additionally, participants reported on their in-role (Abramis 1994) and extra-role performance (Morrison 1994), as well as their engagement (Schaufeli and Bakker 2003), burnout (Schaufeli and van Dierendonck 1993), and job satisfaction (DeWitte et al. 2001).

Sample 2 measured motivation through an adapted version of the MWMS (Gagné et al. 2015) in which the external (material) regulation was not measured due to the voluntary nature of work and lack of applicability to such work. Volunteers also completed self-report measures of individual level performance, including subscales for task proficiency, adaptively, and proactivity (Griffin et al. 2007), as well as measures of organizational identification (Mael and Tetrick 1992), thriving (Porath et al. 2012) and intention to stay at the organization.

In Sample 3, an adaptation of the Academic Motivation Scale (AMS; Vallerand et al. 1989) was used to capture amotivation, external, introjected, and identified regulations, as well as intrinsic motivation. Self-reported outcome variables included students' perceived competence (Losier et al. 1993) and student engagement (Clifford 1988). Student grade point averages (GPAs) in core subjects (French, mathematics, English, and physical education) were also obtained from school records, and an aggregate GPA computed for each time point. Outcomes were examined at all time points.

Sample 4 participants completed the behavioral regulation in exercise questionnaire (BREQ; Mullan et al. 1997) which measured external, introjected, identified, and intrinsic motivation. Additionally, a single item report of how often participants had engaged in leisure-time physical activity in the past week was measured at each of the two follow-up time points. Motivation scores from the time 1 administration were used, whereas exercise behavior from both timepoints 2 and 3 was used as outcomes in the current study.

In Sample 5 students responded to the revised perceived locus of causality in physical education scale (PLOC-R; Vlachopoulos et al. 2011) which measured amotivation, external regulation, introjected regulation, identified regulation, and intrinsic motivation. Additionally, they reported on their level of vitality (Ryan and Frederick 1997) and completed three subscales measuring psychological need frustration of the Psychological Need Thwarting Scale (Bartholomew et al. 2011). Outcomes were taken from timepoints 2 and 3, whereas motivation scores were collected at time 1.

Sample 6 completed the Sports Motivation Scale-II (Pelletier et al. 2013), which measures six regulation subscales including amotivation, external regulation, introjected regulation, identified regulation, integrated regulation, and intrinsic motivation. The dataset also included the satisfaction with life scale (Diener et al. 1985) and the subjective vitality scale (Ryan and Frederick 1997), which were incorporated as outcomes.

## Analyses

Across all samples measurement models were first estimated and resulting factor scores were saved for further use. In comparison with scale scores (i.e., averages of items on a subscale), factor scores have the advantage of providing a partial control for measurement error by giving more weight to items presenting lower levels of measurement error and better conserving the underlying nature of measurement models (Morin et al. 2016a, b; Skrondal and Laake 2001). This approach represented an ideal method in which all computational models could be estimated and applied under a consistent method. In doing so, we could control for statistical artifacts, specifically measurement error, and isolate the concept of interest, that is, the method by which motivation is aggregated. Outcome variables were estimated independently through standard CFA procedures. All analyses were conducted in Mplus (version 7.3; Muthén and Muthén 2015).

### RAI

The RAI (Grolnick and Ryan 1989) was calculated from motivation subscale factor scores, with weights assigned according to the formula presented in the Introduction. It is important to note that, in prior research, the RAI has typically been calculated through averaged scale scores and not latent factors (or factor scores as is the case in the current study). Hence, RAI scores typically do not control for measurement error and instead compound this error multiplicatively. In calculating the RAI, we control for this error and thus control for measurement error equally across scoring methods and thereby are able to attribute differences between scoring methods to information loss rather than differences in measurement error.

### Higher-order model

Higher-order models were specified with external and introjected first order factors loading onto the higher-order "controlled motivation" factor, and intrinsic and identified first order factors loading onto an "autonomous motivation" higher order factor. When measured, we modeled amotivation as a separate latent factor. In the one samples in which integrated regulation was measured, it was specified to load onto the "autonomous motivation" factor alongside identified regulation and intrinsic motivation. Autonomous and controlled factors were permitted to correlate as is standard procedure. While alternate specifications of the higher-order model are sometimes applied in research, we tested only the ideally specified model in which error is reduced and parameters are not added or removed based upon fit statistics.

### CFA

CFA models were estimated in line with standard independent cluster model (ICM) constraints. As such, item loadings on their a priori motivation factor were freely estimated, cross-loadings constrained to zero, and correlations between latent factors permitted.

### ESEM and B-ESEM

Finally, exploratory structural equation modeling (ESEM) and B-ESEM were conducted on the motivation scales. All ESEM and B-ESEM models were specified with a priori expectations about factor structures (similar to CFA), but simultaneously allowed for the free estimation of all possible cross-loadings, using target rotation (Asparouhov and Muthén 2009). While ESEM models freely estimated correlations between motivation factors (as in CFA), B-ESEM models were specified such that the S-factors were orthogonal as the G-factor captures variance that is common to all motivation items (see Morin et al. 2016a, b).

### Regression

The second step involved conducting a series of SEM analyses in which available outcomes in each sample were regressed onto the saved motivation factor scores from each scoring method. For example, in each sample the standard CFA motivation factors were first used to predict outcomes, before the analysis was re-run with the RAI predicting the same outcomes instead of the CFA factors. The resulting explained variance ($R^2$) was recorded for each of these analyses (Tables S3, S5, S7, S9, S11, and S13), and the individual regression coefficients associated with each motivation type recorded in Tables S4, S6, S8, S10, S12, and S14. In order to interpret these results, we first compared the explained variance within samples. This was done by converting the estimated $R^2$ into a proportion representing the variance explained in outcomes by each computational method when compared to the CFA model (i.e., $R_{comparison}^2/R_{CFA}^2$). The CFA model was selected as the comparison group as, of the commonly applied methods, it most closely represents the multidimensional nature of motivation which characterizes SDT. Because results were compared within datasets, for all comparative results, factors such as age, gender, scale used, nationality, and outcomes examined were held constant. This resulted in six sets of results (see Tables S15-S20). These results were then averaged across samples in order to increase interpretability and generalizability of results (Table 3). Additionally, in line with standard procedures, the RMSEA, CFI, and TLI fit statistics were used as primary indicator of model fit. Additionally, information criteria were used to compare competing scoring methods, namely the Akaïke Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Adjusted Bayesian Information Criterion (ABIC), and the Consistent Akaïke Information Criterion (CAIC). Given that the RMSEA and TLI statistics are adjusted for parsimony (Marsh et al. 2009), these criteria provide estimates of goodness of model fit for competing models and apply penalties for added complexity, and in doing so aim to avoid over-fitting models with unnecessary parameters. Heuristically, these criteria indicate whether the added explanatory power of a model warrants the added complexity.

## Results

Descriptive statistics of variables in all samples are provided in Table S2 in the supplementary materials. The fit statistics for the measurement models (Table 1), unsurprisingly, show that B-ESEM measurement models consistently provided the best model fit, closely followed by ESEM. CFA models fit less well (and below generally accepted standards in four of six samples) than either B-ESEM or ESEM, but better than higher-order models across the samples. It is worth noting that B-ESEM models did not converge in sample four and neither ESEM nor B-ESEM models converged in sample six.

**Table 1** Measurement models fit statistics

| Scoring Method | $\chi^2$ | df | #fp | RMSEA (90% CI) | CFI | TLI | AIC | BIC | CAIC | ABIC |
|---|---|---|---|---|---|---|---|---|---|---|
| Sample 1 (workplace) n = 621 | | | | | | | | | | |
| CFA | 504.839* | 137 | 72 | 0.066 (0.060; 0.072) | 0.898 | 0.873 | 36,671.75 | 36,990.80 | 37,062.80 | 36,762.21 |
| Higher-order | 1095.779* | 148 | 61 | 0.102 (0.096; 0.107) | 0.737 | 0.696 | 37,410.43 | 37,680.74 | 37,741.74 | 37,487.08 |
| ESEM | 148.728* | 72 | 137 | 0.041 (0.032; 0.051) | 0.979 | 0.949 | 36,349.67 | 36,956.76 | 37,093.76 | 36,521.80 |
| BESEM | 95.397* | 59 | 150 | 0.032 (0.019; 0.043) | 0.990 | 0.971 | 36,316.48 | 36,981.18 | 37,131.18 | 36,504.95 |
| Sample 2 (volunteer) n = 210 | | | | | | | | | | |
| CFA | 436.750* | 194 | 81 | 0.077 (0.068; 0.087) | 0.857 | 0.830 | 11,971.42 | 12,242.54 | 12,323.54 | 11,985.88 |
| Higher-order | 473.864* | 204 | 71 | 0.079 (0.070; 0.089) | 0.841 | 0.820 | 11,989.41 | 12,227.06 | 12,298.06 | 12,002.09 |
| ESEM | 256.864* | 114 | 161 | 0.077 (0.065; 0.090) | 0.916 | 0.829 | 11,879.24 | 12,418.13 | 12,579.13 | 11,907.99 |
| BESEM | 166.810* | 98 | 177 | 0.058 (0.042; 0.073) | 0.959 | 0.904 | 11,850.63 | 12,443.07 | 12,620.07 | 11,882.23 |
| Sample 3 (education) n = 945 | | | | | | | | | | |
| CFA | 671.775* | 160 | 18 | 0.058 (0.054; 0.063) | 0.924 | 0.910 | 65,347.57 | 65,686.86 | 65,704.86 | 65,464.55 |
| Higher-order | 1421.329* | 167 | 63 | 0.089 (0.085; 0.093) | 0.814 | 0.788 | 66,662.81 | 66,968.44 | 67,031.44 | 66,768.35 |
| ESEM | 340.651* | 100 | 130 | 0.050 (0.045; 0.056) | 0.964 | 0.932 | 65,361.97 | 65,992.62 | 66,122.62 | 65,579.75 |
| BESEM | 264.601* | 85 | 145 | 0.047 (0.041; 0.054) | 0.973 | 0.940 | 65,274.34 | 65,977.76 | 66,122.76 | 65,517.25 |
| Sample 4 (exercise) n = 233 | | | | | | | | | | |
| CFA | 148.870* | 84 | 51 | 0.058 (0.042; 0.072) | 0.948 | 0.935 | 11,984.61 | 12,160.61 | 12,211.61 | 11,998.96 |
| Higher-order | 166.488* | 87 | 48 | 0.063 (0.048; 0.077) | 0.936 | 0.923 | 11,998.19 | 12,163.84 | 12,211.84 | 12,011.71 |
| ESEM | 61.510* | 51 | 84 | 0.030 (0.000; 0.054) | 0.992 | 0.983 | 11,947.64 | 12,237.52 | 12,321.52 | 11,971.29 |
| BESEM | – | – | – | – | – | – | – | – | – | – |
| Sample 5 (physical education) n = 416 | | | | | | | | | | |
| CFA | 455.490* | 160 | 70 | 0.068 (0.061; 0.076) | 0.874 | 0.851 | 29,462.11 | 29,740.99 | 29,810.99 | 29,518.87 |
| Higher-order | 627.435* | 167 | 63 | 0.081 (0.075; 0.088) | 0.811 | 0.786 | 31,071.10 | 31,325.03 | 31,388.03 | 31,125.12 |
| ESEM | 264.592* | 100 | 130 | 0.063 (0.054; 0.072) | 0.933 | 0.872 | 30,703.94 | 31,227.93 | 31,357.93 | 30,815.40 |
| BESEM | 199.240* | 85 | 145 | 0.057 (0.047; 0.067) | 0.953 | 0.895 | 30,651.23 | 31,235.68 | 31,380.68 | 30,775.56 |
| Sample 6 (sport) n = 241 | | | | | | | | | | |
| CFA | 296.625* | 120 | 69 | 0.078 (0.067; 0.089) | 0.889 | 0.858 | 13,132.35 | 13,372.80 | 13,441.80 | 13,154.08 |
| Higher-order | 683.091* | 132 | 57 | 0.132 (0.122; 0.141) | 0.652 | 0.597 | 13,470.47 | 13,669.10 | 13,726.10 | 13,488.42 |
| ESEM | – | – | – | – | – | – | – | – | – | – |
| BESEM | – | – | – | – | – | – | – | – | – | – |

*CFA* confirmatory factor analyses, *ESEM* exploratory structural equation modelling, *BESEM* bifactor-exploratory structural equation modelling, $\chi^2$ robust chi-square test of exact fit, *df* degrees of freedom, *#fp* number of free parameters, *CFI* comparative fit index, *TLI* Tucker–Lewis index, *RMSEA* root mean square error of approximation, *90% CI* 90% confidence interval, *AIC* Akaïke Information Criteria, *CAIC* constant AIC, *BIC* Bayesian Information Criteria, *ABIC* sample-size adjusted BIC

Higher-order models often required additional minor parameter restrictions to force convergence on suitable solutions

Model fit indicators that account for parsimony (i.e. RMSEA and TLI) likewise suggest ESEM and B-ESEM models are not needlessly complex. CFA models were more parsimonious when compared against higher-order models, though were did not fit as well as ESEM and B-ESEM models according to these indicators. When examining parsimony of regression models through the information criteria (AIC, BIC, CAIC and ABIC; Table 2), results were inconclusive as each of the major scoring methods (i.e. CFA, higher order, and RAI) demonstrated superior model fit in different data sets. When considering parsimony of measurement and regression models together, results demonstrate that the amount of information in a model was directly comparable to the models' complexity such that more simplified models represented less construct-relevant information.

Results of regression analyses ($R^2$) applying the different scoring methods, as well as regression weights associated with each regulation type are presented in Tables S3-S14.

**Table 2** Regression model fit statistics by sample

| Scoring method | #fp | LL | Scaling | AIC | BIC | CAIC | ABIC | ΔAIC | ΔBIC | ΔCAIC | ΔABIC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample 1 (workplace) | | | | | | | | | | | |
| CFA | 50 | – 3416.68 | 1.337 | 6933.351 | 7154.918 | 7204.918 | 6996.176 | | | | |
| Higher-order | 30 | – 3546.33 | 1.352 | 7152.650 | 7285.590 | 7315.590 | 7190.345 | 219.299 | 130.672 | 110.672 | 194.169 |
| RAI | 25 | – 3581.37 | 1.346 | 7212.735 | 7323.518 | 7348.518 | 7244.147 | 279.384 | 168.600 | 143.600 | 247.971 |
| ESEM | 50 | – 3414.75 | 1.334 | 6929.506 | 7151.072 | 7201.072 | 6992.330 | – 3.845 | – 3.846 | – 3.846 | – 3.846 |
| BESEM | 55 | – 3412.74 | 1.315 | 6935.479 | 7179.202 | 7234.202 | 7004.586 | 2.128 | 24.284 | 29.284 | 8.410 |
| Sample 2 (volunteer) | | | | | | | | | | | |
| CFA | 99 | – 2055.436 | 1.1713 | 4308.872 | 4640.236 | 4739.236 | 4326.547 | | | | |
| Higher-order | 81 | – 2069.239 | 1.2086 | 4300.478 | 4571.593 | 4652.593 | 4314.939 | – 8.394 | – 68.643 | – 86.643 | – 11.608 |
| RAI | 63 | – 2102.734 | 1.2799 | 4331.468 | 4542.336 | 4605.336 | 4342.716 | 30.990 | – 29.257 | – 47.257 | 27.777 |
| ESEM | 99 | – 2053.940 | 1.1612 | 4305.881 | 4637.244 | 4736.244 | 4323.555 | – 16.555 | 103.940 | 139.940 | – 10.129 |
| BESEM | 108 | – 2050.188 | 1.1362 | 4316.377 | 4677.864 | 4785.864 | 4335.658 | 10.496 | 40.620 | 49.620 | 12.103 |
| Sample 3 (education) | | | | | | | | | | | |
| CFA | 150 | – 23,158.7 | 1.162 | 46,617.314 | 47,344.992 | 47,494.992 | 46,868.601 | | | | |
| Higher-order | 114 | – 23,245 | 1.186 | 46,718.092 | 47,271.127 | 47,385.127 | 46,909.070 | 100.778 | – 73.865 | – 109.865 | 40.469 |
| RAI | 102 | – 23,315.6 | 1.212 | 46,835.180 | 47,330.001 | 47,432.001 | 47,006.055 | 217.866 | – 14.991 | – 62.991 | 137.454 |
| ESEM | 150 | – 23,159.9 | 1.165 | 46,619.790 | 47,347.467 | 47,497.467 | 46,871.077 | 2.476 | 2.475 | 2.475 | 2.476 |
| BESEM | 162 | – 23,154.2 | 1.161 | 46,632.468 | 47,418.360 | 47,580.360 | 46,903.857 | 15.154 | 73.368 | 85.368 | 35.256 |
| Sample 4 (exercise) | | | | | | | | | | | |
| CFA | 13 | – 2416.79 | 0.914 | 4859.584 | 4904.447 | 4917.447 | 4863.244 | | | | |
| Higher-order | 9 | – 2420.42 | 0.845 | 4858.835 | 4889.894 | 4898.894 | 4861.369 | – .749 | – 14.553 | – 18.553 | – 1.875 |
| RAI | 7 | – 2421.86 | 0.814 | 4857.718 | 4881.875 | 4888.875 | 4859.689 | – 1.866 | – 22.572 | – 28.572 | – 3.555 |
| ESEM | 13 | – 2416.08 | 0.918 | 4858.156 | 4903.019 | 4916.019 | 4861.816 | – 1.428 | – 1.428 | – 1.428 | – 1.428 |
| BESEM | – | – | – | – | – | – | – | – | – | – | – |
| Sample 5 (physical education) | | | | | | | | | | | |
| CFA | 115 | – 3797.49 | 1.303 | 7824.988 | 8288.517 | 8403.517 | 7923.592 | | | | |
| Higher-order | 85 | – 3830.06 | 1.395 | 7830.122 | 8172.730 | 8257.730 | 7903.003 | 5.134 | – 115.787 | – 145.787 | – 20.589 |
| RAI | 75 | – 3842.27 | 1.453 | 7834.532 | 8136.834 | 8211.834 | 7898.840 | 9.544 | – 151.683 | – 191.683 | – 24.752 |
| ESEM | 115 | – 3797.44 | 1.304 | 7824.881 | 8288.409 | 8403.409 | 7923.485 | – 0.107 | – 0.108 | – 0.108 | – 0.107 |
| BESEM | 125 | – 3791.45 | 1.282 | 7832.897 | 8336.732 | 8461.732 | 7940.075 | 7.909 | 48.215 | 58.215 | 16.483 |
| Sample 6 (sport) | | | | | | | | | | | |
| CFA | 17 | – 578.893 | 1.1167 | 1191.785 | 1251.027 | 1268.027 | 1197.140 | | | | |
| Higher-order | 9 | – 583.594 | 1.1868 | 1185.189 | 1216.552 | 1225.552 | 1188.024 | – 6.596 | – 34.475 | – 42.475 | – 9.116 |
| RAI | 7 | – 587.192 | 1.1421 | 1188.383 | 1212.777 | 1219.777 | 1190.588 | – 3.402 | – 38.250 | – 48.250 | – 6.552 |
| ESEM | – | – | – | – | – | – | – | – | – | – | – |
| BESEM | – | – | – | – | – | – | – | – | – | – | – |

All change statistics are relative to CFA models; *LL* model log likelihood, *#fp* number of free parameters, *Scaling* scaling factor associated with MLR log likelihood estimates, *AIC* Akaïke Information Criteria, *CAIC* constant AIC, *BIC* Bayesian Information Criteria, *ABIC* sample-size adjusted BIC

Table S15-S20 present this information when transformed into the proportion of variance accounted for by each scoring method when compared against CFA within each dataset. These results were then averaged over all samples in order to calculate a score indicating the overall effectiveness of each computational model in predicting outcomes. Specifically, Table 3 presents the average proportion of variance accounted for, across all outcomes and all samples, for each computational method as well as the associated 95% confidence intervals.

Results indicate that the CFA solution was consistently among the most predictive models across samples with explained variance ranging from 0.015 (exercise behavior) to 0.563 (burnout) across outcomes ($M_R^2 = 0.188$). B-ESEM and ESEM models typically performed as well as CFA, accounting for between 101 and 94% respectively of the variance in the outcomes compared against CFA models ($M_R^2 = 0.203$ and 0.181). All other methods were found to fall well below the CFA model. Specifically, higher-order and RAI methods were found to explain 75% ($M_R^2 = 0.143$) and 66% ($M_R^2 = 0.134$) as much variance as models utilizing CFA, respectively.

Interestingly, the average standard deviations of explained variance associated with the RAI scoring method was substantially higher than that of higher-order and CFA models (see Table 3). This indicates that the RAI may predict outcomes less consistently and may be more prone to high (or low) effect sizes in certain instances. This is evident in Sample 3, for example, where the RAI ranged from being among the strongest method in predicting student GPA to the weakest predictor of engagement.

## Discussion

This study reviewed and compared different scoring methods of motivational constructs proposed by SDT to determine which has greater statistical soundness including parsimony, and predictive validity. Beginning with model fit, B-ESEM fit the data best, followed by ESEM, CFA, and then higher-order models. Although more complex models such as B-ESEM and ESEM are expected to display greater model fit due to the increased parameterization, it is interesting to note that CFA out-performed higher-order models. Notably, the RAI is not a model-based estimation procedure and therefore cannot be compared to the other methods based on fit indices. Indicators of parsimony favored ESEM and B-ESEM measurement models, though more broadly, the amount of information captured by a model was approximately proportional to its complexity. This suggests that while simplified models do exclude construct-relevant information, it is not beyond expectation. Finally, examination of variance in outcomes showed a clear division with multidimensional scoring methods (CFA, ESEM, and B-ESEM) capable of predicting approximately 30–40% more variance than more simplified methods (i.e. higher-order models and RAI). In all, these results show that multidimensional scoring methods more accurately and fully represent motivation within SDT and should be preferred over simplified models when theoretically relevant.

It is also worth noting that several instance of lost information can be seen when examining regression results closely (Tables S3–S14). Sample 2 demonstrates the potential positive influence of introjection with resulting indicating a positive effect of this motivation type on workplace adaptivity, proactivity, and intention to stay at the organization. Categorizing introjected regulation as a "controlled" form of motivation and weighting it negatively in the RAI calculation excluded these effects. We also show several instances in which identified regulation appears a more important predictor than intrinsic motivation. For example, both Samples 3 and 6 indicate that the predictive capability of identified regulation when predicting GPA and life satisfaction (Tables S8 and S14). Additionally, as demonstrated in Sample 3, external regulation may play an increasingly negative influence when predicting GPA over the three measured timepoints, whereas intrinsic motivation appeared to become an increasingly less important predictor (Table S8). Each of these examples are theoretically consistent representations of information captured through multidimensional methods that is excluded by more simplified methods.

Comparing CFA, ESEM and B-ESEM, it should be noted that ESEM has demonstrated the ability to reduce artificially inflated inter-factor correlations compared to CFA (Asparouhov and Muthen 2009; Guay et al. 2015), which directly reduces the degree of multicollinearity. ESEM thus results in more accurate partitioning of explained variance between

**Table 3** Percentage of variance accounted by each scoring method in comparison to CFA operationalization

| Descriptive statistics | CFA | Higher-order | RAI | ESEM | Bifactor ESEM |
|---|---|---|---|---|---|
| Mean | 100.00 | 74.97 | 66.53 | 94.67 | 101.71 |
| SD | 0.00 | 12.07 | 22.52 | 11.30 | 4.22 |
| 95% CI upper | 100.00 | 78.97 | 73.99 | 98.41 | 103.19 |
| 95% CI lower | 100.00 | 70.97 | 59.06 | 90.92 | 100.22 |

All numbers are percentages

predictors, which overcomes one of the main criticisms of traditional CFA approaches. B-ESEM fares even better in this respect as all factors in this model are uncorrelated (with commonalities absorbed into the G-factor), eliminating multicollinearity entirely. While this partitioning of variance into orthogonal units is a strength in reducing multicollinearity, it also has important implications for the interpretation of regulation factors. Specifically, B-ESEM involve a different interpretation of motivation as it separates what is common across the motivational regulations into a general factor that has been interpreted as representing the degree of self-determination (as judged by the predictable pattern of factor loadings, Howard et al. 2018; Litalien et al. 2017) from the unique motivational characteristics that differentiate the regulations. While in CFA and ESEM solutions, the intrinsic motivation factor, for example, represents both a high level of self-determined motivation as well as a regulation-specific element (e.g. enjoyment), in B-ESEM, these characteristics are separated into a component representing the degree of self-determined (modeled by the G-factor) and a component purely representing "joy" or "interest" (modeled in the S-factor).

A clear example of note relating to this method of partitioning variance is seen in the regression results of Sample 5 (Table S12). While the ESEM methodology indicates no significant relationship between external regulation and any outcomes, the B-ESEM solution offers a more nuanced perspective in which external regulation significantly predicts need frustration, while the general factor of self-determination has a counteracting negative influence. This clearly highlights that while self-determination reduces need frustration, external incentives increase it. When external regulation is defined as a combination of self-determination and external contingencies, as it is in CFA and ESEM solutions, these effects are entirely obscured. We argue that this partitioning deserves further theoretical and empirical consideration in SDT research to verify initial results (Howard et al. 2018; Litalien et al. 2017), particularly in respect to distinguishing and interpreting the contribution of self-determined motivation versus other qualities of the regulations (e.g., self-esteem contingencies, meaningfulness, enjoyment) in predicting outcomes, and what antecedents are likely to affect each of them. However, both ESEM and B-ESEM solutions have the tendency to fail to converge in smaller data sets. This is noticed in two of the current samples and indicates that these methods will not be a viable option for all studies. Therefore, researchers may need to plan for larger sample sizes to use these methods or, alternatively, opt for CFA when necessary.

As such, it appears that simplified methods such as the RAI and higher-order models are useful for examining specific hypotheses regarding only the degree of self-determination, they will not capture regulation specific effects. While CFA is a somewhat suitable solution, ESEM demonstrates some notable advantages and is therefore preferable to CFA when possible. B-ESEM is a statistically strong and theoretically consistent approach to modeling motivation which demonstrates further advantages over either ESEM or CFA, yet further validation work is strongly encourage to more clearly define the nature and roles of the resulting factors.

## Implications for theory, research and practice

First regarding theory, while each scoring method is supported in some way by SDT, application of these methods has notably different theoretical implications. For example, applying the RAI tells the researcher if greater (or lesser) levels of self-determination are significantly associated with a covariate. However, it does not consider the regulation factors themselves, excluding for example, the possibility to test whether identified regulation (with its focus on meaningfulness) or introjection (with its focus on ego-involvement) plays a role. Higher-order models suffer from this same issue. Multidimensional operationalizations, in contrast, contribute theoretical knowledge not available in simplified scoring methods. Specifically, they take into account the particularities of each regulation, which is important as these have shown to be significant in predicting outcomes in the health, education, and work domains (Burton et al. 2006; Gagné et al. 2015; Ng et al. 2012). B-ESEM is particularly noteworthy in this regard as it separates what is common versus specific across the motivational regulations, and therefore may be the most theoretically consistent approach to modeling motivation. These more complex multidimensional scoring methods will be necessary for the future theoretical development of SDT. These issues are theoretically important because selection of a scoring method will dictate the necessary interpretation of results. As each scoring method conceptualizes different components of motivation, researchers must be aware of these implications, and utilize scoring method appropriately to align with research questions in order to be effective.

For empirical research and practice, our results indicate that multidimensional scoring methods should be favored when possible in terms of conforming to theoretical tenets of SDT, statistical soundness, and variance explained in outcomes. Applying these more complex methods may mean dealing with more motivation variables than would be required by simplified methods and, in the case of CFA, additionally dealing with the multicollinearity that comes along with this. One method which largely resolves these issues is relative weights analysis (RWA), which is most easily conducted through a web-based interface which produces syntax based upon imputed information (Tonidandel and LeBreton 2015). This method was designed to remove multicollinearity from correlated predictors and estimate the proportion of explained variance each individual predictor is capable of accounting for (Tonidandel and LeBreton 2011, 2015). As such, using RWA, in conjunction with CFA-derived motivation factors, will clearly identify which regulation types are influencing an outcome (and which are not) and indicate the relative contribution of each. Using a multidimensional approach accompanied by RWA may provide a relatively simple and yet highly effective way forward for many areas of motivation research. For example, among organizational scholars there has been ongoing disagreement concerning the role of external regulation and its ability to predict performance, with some arguing for a positive relationship (Gerhart and Fang 2014, 2015) and others arguing for a null or even negative influence (Gagné and Forest 2008). While an undoubtedly complex topic of study, the role of multicollinearity in shrouding these results remains unclear. The procedures recommended here, specifically, ESEM, CFA accompanied by RWA, or bifactor models will help in solving these issues. While complex, the advantages of not only increasing predictive capability but also modelling motivation more comprehensively (thereby capturing regulation specific effects) is something to aspire to as SDT continues to develop.

In addition to being theoretically meaningful, understanding the contribution of each of the motivation types can have practical implications in the design of interventions, and may help practitioners decide whether they should, for example, direct their efforts towards making activities more interesting and enjoyable (i.e. intrinsically motivating), more meaningful (identified regulation), or whether externally administered rewards (external regulation) are necessary in order to obtain desired outcomes such as continued effort and performance (Gagné et al. 2015; Ng et al. 2012).

### Limitations and future directions

We first acknowledge that the development and application of scoring methods is more complex than well-fitting models and increased amounts of explainable variance. Accordingly, we cannot give a universal recommendation as the scoring method selected will depend on theoretical alignment between method and the hypotheses to be tested. However, we highlight meaningful theoretical implications and show empirical differences between various scoring methods and recommend future research account for these issues. Further practical considerations are also worth noting, for example, the expected sample size and potential restrictions on number of items measured (e.g. experience sampling methodologies and experimental research). While such considerations may result in the use of composite measures of motivation, it must however be acknowledged that such tests are incomplete representations of SDT (Howard et al. 2018; Litalien et al. 2017) and therefore carry demonstrable limitations.

A further possible limitation of the current study is the use of factor scores in the predictive models rather than fully latent models. The decision to use factor scores has moved the results one step away from typical research but has controlled for divergences between methods being either model-based (i.e. SEM) or not (e.g., RAI), and thereby standardized one element of operationalization to enable more direct comparisons. While this ideally suited the current research questions, it is acknowledged that research using fully latent models could obtain slightly different results, likely favoring model-based methods that account for measurement error more fully.

Future study designs could also seek to more clearly distinguish ESEM and B-ESEM solutions and offer more concrete examples of when one method would be considered ideal over the other. While recent evidence is mounting to support the B-ESEM solution (Howard et al. 2018; Litalien et al. 2017), several issues remain to be resolved including the potential meaning of a general factors shifting depending on included subscales, and greater understanding of factors which predict each element within bifactor models. Future research designed to validate this factor structure is recommended. Finally, the current examination of different scoring methods was limited to SDT. We would encourage future research to examine multidimensional constructs within other theoretical frameworks and the implications of using subscales compared to more parsimonious methods. Our expectations are that such endeavors would come to the same list of advantages and disadvantages which would lead to similar recommendations. However, future research is needed to test the generalizability of our results.

## Conclusion

Through testing and comparing multiple computational methods for motivational scales derived from SDT, it is demonstrated that multidimensional scoring methods more accurately and comprehensively represent motivation as evidenced by the greater variance explained in outcomes. While we acknowledge that simplified scoring methods will be suitable in some research contexts examining only the degree of self-determination, it is important to consider the implications of such practices and interpret results accordingly. It is recommended that multidimensional methods (in combination with relative weights analysis when necessary) be used when possible in order to maximize the accuracy and predictive power of SDT's motivational conceptualization, and to capture important regulation-specific effects.

**Acknowledgements** Part of this research was funded through an Australian Research Council grant awarded to the second author (LP150100417).

## References


Abramis, D. J. (1994). Relationship of job stressors to job performance: Linear or inverted-u? *Psychological Reports, 75*, 547–558.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*(3), 397–438.

Bartholomew, K. J., Ntoumanis, N., Ryan, R. M., & Thøgersen-Ntoumani, C. (2011). Psychological need thwarting in the sport context: Assessing the darker side of athletic experience. *Journal of Sport and Exercise Psychology, 33*(1), 75–102.

Bono, J. E., & Judge, T. A. (2003). Self-concordance at work: Understanding the motivational effects of transformational leaders. *Academy of Management Journal, 46*, 554–571.

Burton, K. D., Lydon, J. E., D'Alessandro, D. U., & Koestner, R. (2006). The differential effects of intrinsic and identified motivation on well-being and performance: Prospective, experimental, and implicit approaches to self-determination theory. *Journal of Personality and Social Psychology, 91*(4), 750.

Chatzisarantis, N. L., Hagger, M. S., Biddle, S. J., Smith, B., & Wang, J. C. (2003). A meta-analysis of perceived locus of causality in exercise, sport, and physical education contexts. *Journal of Sport and Exercise Psychology, 25*(3), 284–306.

Chemolli, E., & Gagné, M. (2014). Evidence against the continuum structure underlying motivation measures derived from self-determination theory. *Psychological Assessment, 26*(2), 575.

Clifford, M. M. (1988). Failure tolerance and academic risk-taking in ten-to twelve-year-old students. *British Journal of Educational Psychology, 58*(1), 15–27.

DeWitte, H., Hooge, J., Vandoorne, J., & Glorieux, I. (2001). *Prettig werken in een gezonde samenleving. Omvang, verschillen en gevolgen van arbeidstevredenheid in Vlaanderen*. In Administratie Planning en Statistiek (Ed.), Vlaanderen gepeild! De Vlaamse overheid en burgeronderzoek 2001 (pp. 73–98). Brussel: Ministerie van de Vlaamse Gemeenschap

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71–75.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods, 4*, 265–287. https://doi.org/10.1177/109442810143005.

Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., Van den Broeck, A., Aspeli, A. K., … Halvari, H. (2015). The Multidimensional Work Motivation Scale: Validation evidence in seven languages and nine countries. *European Journal of Work and Organizational Psychology*, *24*(2), 178-196.

Gerhart, B., & Fang, M. (2014). Pay for (individual) performance: Issues, claims, evidence and the role of sorting effects. *Human Resource Management Review, 24*(1), 41–52.

Gerhart, B., & Fang, M. (2015). Pay, intrinsic motivation, extrinsic motivation, performance, and creativity in the workplace: Revisiting long-held beliefs. *Annual Review of Organizational Psychology & Organizational Behavior, 2*(1), 489–521.

Gillet, N., Gagné, M., Sauvagère, S., & Fouquereau, E. (2013). The role of supervisor autonomy support, organizational support, and autonomous and controlled motivation in predicting employees' satisfaction and turnover intentions. *European Journal of Work and Organizational Psychology, 22*(4), 450–460.

Griffin, M. A., Neal, A., & Parker, S. K. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal, 50*(2), 327–347.

Grolnick, W. S., & Ryan, R. M. (1989). Parent styles associated with children's self-regulation and competence in school. *Journal of Educational Psychology, 81*(2), 143.

Guay, F., Mageau, G. A., & Vallerand, R. J. (2003). On the hierarchical structure of self-determined motivation: A test of top-down, bottomup, reciprocal, and horizontal effects. *Personality and Social Psychology Bulletin, 29*, 992–1004.

Guay, F., Morin, A. J., Litalien, D., Valois, P., & Vallerand, R. J. (2015). Application of exploratory structural equation modeling to evaluate the academic motivation scale. *The Journal of Experimental Education, 83*(1), 51–82.

Howard, J. L., Gagné, M., & Bureau, J. S. (2017). Testing a continuum structure of self-determined motivation: A meta-analysis. *Psychological Bulletin, 143*(12), 1346.

Howard, J. L., Gagné, M., Morin, A. J., & Forest, J. (2018). Using bifactor exploratory structural equation modeling to test for a continuum structure of motivation. *Journal of Management, 44*, 2638–2664.

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organization Behavior and Human Performance, 27*, 443–463.

Li, F. (1999). The exercise motivation scale: Its multifaceted structure and construct validity. *Journal of Applied Sport Psychology, 11*(1), 97–115.

Litalien, D., Morin, A. J., Gagné, M., Vallerand, R. J., Losier, G. F., & Ryan, R. M. (2017). Evidence of a continuum structure of academic self-determination: A two-study test using a bifactor-ESEM representation of academic motivation. *Contemporary Educational Psychology, 51*, 67–82.

Losier, G. F., & Koestner, R. (1999). Intrinsic versus identified regulation in distinct political campaigns: The consequences of following politics for pleasure versus personal meaningfulness. *Personality and Social Psychology Bulletin, 25*, 287–298.

Losier, G. F., Vallerand, R. J., & Blais, M. R. (1993). Construction et validation de l'Échelle des Perceptions de Compétence dans les Domaines de Vie (EPCDV). *Science et Comportement, 23*, 1–16.

Mael, F. A., & Tetrick, L. E. (1992). Identifying organizational identification. *Educational and Psychological Measurement, 52*(4), 813–824.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*, 85–110.

Morin, A. J. S., Arens, A., & Marsh, H. (2016a). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling, 23*, 116–139.

Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling: An introduction. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 395–436). Greenwich, CT: IAP.

Morin, A. J. S., Meyer, J. P., Creusier, J., & Biétry, F. (2016b). Multiple-group analysis of similarity in latent profile solutions. *Organizational Research Methods, 19*, 231–254.

Morrison, E. W. (1994). Role definitions and organisational citizenship behaviour: The importance of the employee's perspective. *Academy of Management Journal, 37*, 1543–1567.

Mullan, E., Markland, D., & Ingledew, D. K. (1997). A graded conceptualisation of self-determination in the regulation of exercise behaviour: Development of a measure using confirmatory factor analytic procedures. *Personality and Individual Differences, 23*(5), 745–752.

Muthén, L. K., & Muthén, B. (2015). *Mplus: The comprehensive modelling program for applied researchers: user's guide, 5*. Los Angeles, CA: Muthen & Muthen.

Ng, J. Y., Ntoumanis, N., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Duda, J. L., et al. (2012). Self-determination theory applied to health contexts: A meta-analysis. *Perspectives on Psychological Science, 7*(4), 325–340.

Pelletier, L. G., Fortier, M. S., Vallerand, R. J., & Briere, N. M. (2001). Associations among perceived autonomy support, forms of self-regulation, and persistence: A prospective study. *Motivation and Emotion, 25*(4), 279–306.

Pelletier, L. G., Rocchi, M. A., Vallerand, R. J., Deci, E. L., & Ryan, R. M. (2013). Validation of the revised sport motivation scale (SMS-II). *Psychology of Sport and Exercise, 14*(3), 329–341.

Pelletier, L. G., Tuson, K. M., Green-Demers, I., Noels, K., & Beaton, A. M. (1998). Why are you doing things for the environment? The motivation toward the environment scale (mtes) 1. *Journal of Applied Social Psychology, 28*(5), 437–468.

Phillips, L. A., & Johnson, M. A. (2018). Interdependent effects of autonomous and controlled regulation on exercise behavior. *Personality and Social Psychology Bulletin, 44*(1), 49–62.

Porath, C., Spreitzer, G., Gibson, C., & Garnett, F. G. (2012). Thriving at work: Toward its measurement, construct validation, and theoretical refinement. *Journal of Organizational Behavior, 33*(2), 250–275.

Reeve, J. (2016). A grand theory of motivation: Why not? *Motivation and Emotion, 40*(1), 31–35.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*(5), 749–761.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. New York: Guilford Publications.

Ryan, R. M., & Frederick, C. (1997). On energy, personality, and health: Subjective vitality as a dynamic reflection of well-being. *Journal of Personality, 65*(3), 529–565.

Schaufeli, W. B., & Bakker, A. B. (2003). *UWES-Utrecht Work Engagement Scale: Test manual.* Unpublished manuscript, Department of Psychology, Utrecht University.

Schaufeli, W. B., & van Dierendonck, D. (1993). The construct validity of two burnout measures. *Journal of Organizational Behavior, 14*, 631–647.

Sheldon, K. M., Osin, E. N., Gordeeva, T. O., Suchkov, D. D., & Sychev, O. A. (2017). Evaluating the dimensionality of self-determination theory's relative autonomy continuum. *Personality and Social Psychology Bulletin, 43*(9), 1215–1238.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika, 66*, 563–576.

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology, 26*(1), 1–9.

Tonidandel, S., & LeBreton, J. M. (2015). RWA web: A free, comprehensive, web-based, and user-friendly tool for relative weight analyses. *Journal of Business and Psychology, 30*(2), 207–216.

Vasconcellos, D., Parker, P. D., Hilland, T., Cinelli, R., Owen, K. B., Kapsal, N., … Lonsdale, C. (2019). Self-determination theory applied to physical education: A systematic review and meta-analysis. *Journal of Educational Psychology*. 10.1037/edu0000420.

Vallerand, R. J., Blais, M. R., Brière, N. M., & Pelletier, L. G. (1989). Construction et validation de l'échelle de motivation en éducation (EME). *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement, 21*(3), 323–349.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senécal, C., & Vallieres, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement, 52*, 1003–1017.

Vlachopoulos, S. P., Katartzi, E. S., Kontou, M. G., Moustaka, F. C., & Goudas, M. (2011). The revised perceived locus of causality in physical education scale: Psychometric evaluation among youth. *Psychology of Sport and Exercise, 12*(6), 583–592.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.