UNIVERSITY of CALIFORNIA PRESS | Collabra: Psychology

Bieleke, M., Stähler, J., Wolff, W., & Schüler, J. (2025). Development and Validation of the Value of Physical Effort (VoPE) Scale. *Collabra: Psychology*, *11*(1). https://doi.org/10.1525/collabra.140736

Personality Psychology

Development and Validation of the Value of Physical Effort (VoPE) Scale

Maik Bieleke[1][a], Johanna Stähler[1], Wanja Wolff[2], Julia Schüler[1]

[1] Department of Sport Science, University of Konstanz, Konstanz, Germany, [2] Dynamics of Human Performance Regulation Laboratory, Department of Movement Science, University of Hamburg, Hamburg, Germany

Keywords: Value of Physical Effort (VoPE), sports and exercise, physical activity, networks, forecasting, out-of-sample prediction, machine learning

https://doi.org/10.1525/collabra.140736

Collabra: Psychology

Vol. 11, Issue 1, 2025

Physical effort has instrumental value because it helps people attain their goals. Growing evidence suggests that people might also experience the exertion of effort itself as valuable. To test this idea, we developed and examined the 4-item Value of Physical Effort (VoPE) scale. Across three independent studies (total $N = 1364$), we established the basic psychometric properties of the VoPE scale and showed consistent associations with measures of sports and exercise behavior. In a study with a longitudinal design, the VoPE scale demonstrated test-retest reliability and forecasted physical activity and exercise behavior. Psychometric network analysis and elastic net regression suggest the VoPE scale's potential to cover a unique content area and make novel contributions to the prediction of sports behavior. Taken together, the present research yields empirical evidence for the idea that people value physical effort to varying degrees, and that these differences can be efficiently measured with the VoPE scale. We demonstrated reliability and validity of the VoPE scale across three datasets and its ability to predict self-reported sports and exercise behavior. The VoPE scale might thus extend our understanding of how people allocate (physical) effort.

Introduction

> *Success is dependent on effort.* (Sophocles, ancient Greek writer)

> *Hard work is the price we must pay for success.* (Vince Lombardi, American football coach)

> *Success is the sum of small efforts, repeated day-in and day-out.* (Robert Collier, American author of self-help books)

As the introductory quotes above emphasize, it is a widespread notion that consistent mobilization effort is crucial for achieving one's goals. Consistent with this assertion, a large body of research attests to the instrumentality of effort as a means for success in various domains. On a macro level, for example, higher effort predicts better work performance (Van Iddekinge et al., 2023), better school achievement (Carbonaro, 2005), and general success in life (Moffitt et al., 2011). On a micro level, higher effort is linked to better performance in psychological tests (Silm et al., 2020) and to trial-by-trial variations in physical force output (Rewitz et al., 2024). As these findings highlight, cognitive and physical effort are crucial for goal attainment. However, most readers are probably keenly aware of the fact that trying harder (i.e., investing more effort) does not always lead to better performance (Halperin & Vigotsky, 2023), and even *if* effort is conducive to success, its exertion still feels aversive and people try to avoid it (e.g., David et al., 2022; Kool & Botvinick, 2014; Kurzban, 2016). Indeed, a large body of theoretical and empirical evidence indicates that humans tend to adhere to the "principle of least effort" (Zipf, 1949) and make effort mobilization contingent on the rewards they are trying to attain (Brehm & Self, 1989; Brinkmann et al., 2021). Importantly, this principle of least effort seems to hold for cognitive effort and physical effort and people generally try to minimize effort in both domains. Thus, while effort is instrumental for success, humans try to mobilize it as little as possible.

But do all people make effort exertion contingent on its instrumentality for goal attainment? Recent work suggests that this might not always be the case. Under *some* conditions *some* people might even mobilize effort because to them effort exertion is rewarding in its own right (Inzlicht et al., 2018; Stähler et al., 2025). Take for example recreational athletes: Every year, millions of people invest substantial resources (e.g., money, time, risk of injury) to

a Corresponding author: Maik Bieleke, Department of Sport Science, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany.

participate in exhausting endurance events with negligible prospects of obtaining tangible rewards in return (Maxcy et al., 2019). And while there surely are non-tangible rewards to justify the effort (e.g., better fitness, social engagement), these people still seem to attach levels of value to the experience of physical effort that are difficult to reconcile with a purely instrumental perspective on effort. Put differently, they might engage in sports and exercise not despite the physical effort it requires but because of it (Wolff, Hirsch, et al., 2021), suggesting that they value effort to some degree. However, valuing effort is not limited to the domain of physical effort and sports. For example, taking pleasure in the exertion of effort has been identified as a potential explanation of Anorexia Nervosa patients' excessive engagement in effortful behaviors that most other people find highly aversive (e.g., restrictive eating; Haynos et al., 2022). Taken together, while people generally try to minimize effort, some people seem to find the exertion of effort not aversive, but valuable and rewarding. This points towards individual differences with respect to how people value (or do not value) effort.

Interestingly, in the cognitive effort domain the role of individual differences in valuing effort has long been acknowledged: The concept of a "need for cognition" (Cacioppo & Petty, 1982) essentially reflects variations in how much people enjoy to engage in cognitively demanding tasks. And meta-analytic evidence supports the "existence of stable individual differences in people's tendency to engage in and enjoy effortful cognitive activity" (Cacioppo et al., 1996, p. 247). To illustrate, when compared to people who score low on a measure of need for cognition, people who value cognitive effort preferred to engage in cognitively demanding leisure time activities, such as writing poetry or reading technical books (Therriault et al., 2015).

As the example of recreational athletes vividly illustrates, it is plausible that people also exhibit stable individual differences with respect to the valuation of physical effort. After all, most people are not recreational athletes but rather struggle to even meet the WHO's recommendations for the amount of weekly physical activity that is needed to maintain one's health – despite widely acknowledged health benefits of exercise (Wang & Ashokan, 2021) and the widespread intention to become more physically active (Rhodes & Dickau, 2012). Thus, while some people seem to treat physical effort as if it was valuable in its own right, many try to avoid and minimize it as much as possible because it feels aversive and costly (Brehm & Self, 1989; Kool & Botvinick, 2014; Kurzban, 2016). Unlike the cognitive domain, however, there is currently no instrument for assessing individual differences in the value of physical effort. This is unfortunate, as such a scale has the potential to shed light on why some people engage more or less than others in physically effortful tasks.

Present Research

We developed the Value of Physical Effort (VoPE) scale across three independent studies. In terms of its basic psychometric properties, we expected the VoPE scale to capture the value of physical effort reliably and as a unidimensional construct (Studies 1 to 3). Regarding its predictive validity, we assumed higher VoPE scores to predict higher levels of self-reported sports and exercise behavior (e.g., exercise volume, experience with sports competitions, physical activity) both cross-sectionally (statistical prediction; Studies 1 and 3) and longitudinally (forecasting; Study 2). Exploratively, we examined whether the VoPE scale contributes meaningfully to existing questionnaires used in sport psychology, and gauged its unique contribution to explaining indicators of sports behavior (Study 3). To maximize the precision of our analyses, we collapsed data across studies where possible and summarized the results with meta-analytic techniques.

Methods

Participants

An overview of participant characteristics in all three studies is provided in Table 1. Participants in Study 1 (N = 264) were recruited at the University of Konstanz, Germany, as part of a survey that included the VoPE scale among other instruments. Participants in Studies 2 (N = 295) and 3 (N = 805) were recruited using Amazon Mturk (requirements: ⩾ 90% approval rate, ⩾ 100 HITs, US citizenship) and received $1 and $2, respectively. To ensure a high quality of responses among the Mturk participants, we used tools like bot captchas and instructional manipulation checks. Participants who did not pass the instructional manipulation checks were disregarded in the analyses (Study 2: 17 participants, Study 3: 61 participants). Moreover, we used unique Mturk IDs to prevent multiple participation. Study 2 was split into two parts that were conducted five to six weeks apart. A subsample of N = 200 participants from the first part was invited to participate in the second part. In Study 3, we included a screening to make sure that participants engaged in at least 30 minutes of physical exercise per week. This was done because some questionnaires could be answered in a meaningful way only with a minimum level of exercise (e.g., instruments pertaining to the reasons for exercising).

Materials and Procedure

We focus on the measures and procedures relevant to addressing our research questions. All materials and the corresponding data can be accessed via OSF (https://osf.io/zk9qr/). The study protocol and measurements were approved by the Ethics Committee at the University of Konstanz (approval IRB25KN003-04/w).

Study 1: VoPE Scale and Sports and Exercise Behavior

Value of Physical Effort (VoPE) scale. In Study 1, the VoPE scale was administered as part of a general survey. To construct the VoPE scale, we adapted items from the Need for Cognition Scale (NfCS; Cacioppo et al., 1984) — a well-established instrument that measures "an individual's tendency to engage in and enjoy effortful cognitive endeavors"

Table 1. Participant characteristics in Studies 1 to 3.

		Study 1	Study 2		Study 3
			Part 1	Part 2	
Sample size	*N*	264	295	197	805
Age	*M* (years) ± SD	27.9 ± 11.7	37.6 ± 9.5	38.2 ± 9.1	40.4 ± 11.3
Gender	female	48.5%	39.0%	36.0%	48.4%
	male	40.8%	60.7%	63.5%	50.7%
	other	0.8%	0.3%	0.5%	0.9%
Exercise volume	*M* (min/w) ± *SD*	261 ± 211	239 ± 188	224 ± 158	306 ± 213
Competition experience	none	36.7%	60.3%	57.9%	55.3%
	recreational	25.8%	25.8%	28.4%	33.8%
	regional	25.4%	9.2%	9.1%	8.9%
	national	8.7%	4.7%	4.6%	2.0%
	international	3.4%	0.0%	0.0%	0.0%
Income	no income	-	0.3%	0.5%	0.8%
	< $5,000	-	2.4%	1.0%	3.11%
	$5,000 to $11,999	-	8.1%	7.1%	5.10%
	$12,000 to $19,999	-	10.5%	9.1%	7.8%
	$20,000 to $39,999	-	24.1%	23.9%	23.6%
	$40,000 to $59,999	-	27.1%	26.4%	22.7%
	$60,000 to $79,999	-	15.3%	16.8%	19.1%
	> $80,000	-	12.2%	15.2%	17.7%
Education	0 to 5 grades	-	0.0%	0.00%	0.0%
	6 to 8 grades	-	0.0%	0.0%	0.1%
	9 to 11 grades	-	0.7%	0.5%	0.1%
	12 grades; High School	-	13.9%	12.7%	9.9%
	12 grades + Training	-	3.4%	3.6%	3.6%
	13 to 14 years	-	24.1%	25.9%	25.5%
	15 to 16 years	-	46.1%	46.2%	44.2%
	17 or more years	-	11.9%	11.2%	16.5%
Employment	Disabled	-	0.3%	0.5%	0.5%
	Homemaker	-	2.0%	1.0%	3.4%
	Retired	-	1.0%	0.5%	2.7%
	Self-employed	-	16.3%	15.7%	12.7%
	Student	-	1.4%	1.0%	1.0%
	Unemployed, not seeking work	-	0.3%	0.5%	0.9%
	Unemployed, seeking work	-	3.1%	2.0%	3.4%
	Working part-time, <35 h/w	-	7.5%	6.6%	8.1%
	Working full-time, >35 h/w	-	68.1%	72.1%	67.5%
Subjective SES	*M* ± *SD*	-	3.6 ± 1.0	3.6 ± 0.9	3.9 ± 1.1

Note. min/w refers to minutes per week. h/w refers to hours per week. Subjective SES refers to subjective socioeconomic status measured on a Likert scale from 1 (least money, education, and respected jobs) to 7 (most money, education, and respected jobs)

(p. 306). We jointly reviewed, discussed, and ranked the 34 NfCS items regarding their suitability for adoption to the domain of sports and exercise. We ended up with a set of ten items that provided the best fit in terms of adoptability (5 positively and 5 negatively framed items) and administered them in all three studies. We translated the VoPE scale from German (Study 1) to English (Studies 2 and 3) with the assistance of AI technology (www.deepl.com; see Table 1 for both language versions). After data collection, we reduced the items to the four items that make up the final VoPE scale (see Table 2).

First, we excluded the five negatively framed items because they formed a separate community in network analyses, likely reflecting a common method artifact that plagues mixed framing scales (Lindwall et al., 2012). The NfCS has also been criticized for introducing artificial factors through its negatively framed items (Zhang et al., 2016), which we aimed to avoid. Second, we removed one of the positively framed item which, in retrospection, referred to an artificial situation that might elicit ambivalent responses: "When I know that a sports activity is going to involve physical effort, I look forward to it, even if I do not know what exactly it involves." Despite these exclusions, all ten items are available in the published dataset on OSF for transparency. Participants indicated their agreement with each item on a 7-point Likert scale (1 = *strongly disagree*, 7 = *strongly agree*), and their responses were averaged into an overall VoPE score.

Exercise volume and competition experience. We used two measures of sports and exercise behavior: First, participants reported the number of exercise sessions per week (frequency) and how long each session typically takes (duration). The exercise volume in minutes per week was computed as volume = frequency x duration. Second, par-

Table 2. The four items of the Value of Physical Effort (VoPE) scale.

No.	English	German
1	I greatly enjoy sports activities that require physical effort.	Ich habe großen Spaß an Aktivitäten, die körperliche Anstrengung erfordern.
2	I prefer physically effortful sports activities to those that can be done without much effort.	Ich bevorzuge körperlich anstrengende Aktivitäten gegenüber sportlichen Aktivitäten, die ohne größere Anstrengung bewältigt werden können.
3	The notion of physically exerting myself in sports appeals to me.	Die Vorstellung, mich beim Sport körperlich anzustrengen, reizt mich.
4	I enjoy physically exerting myself in a sports activity even when it will have no effect on the result.	Es macht mir Spaß, mich bei einer sportlichen Aktivität körperlich anzustrengen, selbst wenn es keinen Einfluss auf das Ergebnis hat.

Note. The items were administered in German in Study 1 and in English in Studies 2 and 3.

ticipants indicated their highest level of experience with sport competitions. They could choose one of the categories (1) no competitions, (2) recreational competitions, (3) regional competitions, (4) national competitions, and (5) international competitions. As the distinction between having no experience at all and having experience seems most significant (e.g., engaging in training) and because rather few participants reported experience with national and international competitions (2.0% – 12.3% of participants; see Table 1), we collapsed answers into a binary variable indicating whether participants had no competition experience (Category 1; 36.7% to 60.3% of participants) or some competition experience (Categories 2 to 5; 39.7% to 63.3% of participants). A similar categorization has been used in related research as well (e.g., Schüler et al., 2023).

Study 2: Test-Retest and Measures of Physical Activity and Exercise

Like in Study 1, participants in Study 2 completed the VoPE scale, indicated their general sports and exercise behavior, and provided demographic information. Five to six weeks later, they completed the VoPE scale again along with two instruments assessing physical activity and exercise behavior.

International Physical Activity Questionnaire (IPAQ). The IPAQ (Craig et al., 2003) measures vigorous physical activity, moderate physical activity, and walking over the past seven days in terms of their frequency (3 items; e.g., "During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, digging, aerobics, or fast bicycling?") and duration (3 items; e.g., "How much time did you usually spend doing vigorous physical activities on one of those days?"). We computed an overall volume for each kind of activity by multiplying frequency and duration. Additionally, the IPAQ assesses the duration of sitting (1 item; "During the last 7 days, how much time did you spend sitting on a week day?"). To get a nuanced picture of physical activity, we analyzed each kind of activity separately rather than relying on an overall score.

Godin-Shephard Leisure-Time Physical Activity Questionnaire (GSLTPAQ). While the IPAQ relates to all kinds of physical activities, the GSLTPAQ (Godin & Shephard, 1985) focuses more specifically on physical exercise at different intensities over the past seven days. It comprises three items to assess how often per week participants engage in strenuous exercise, moderate exercise, and mild exercise. Additionally, it measures how often participants engage in any leisure-time activity long enough to work up a sweat (options: often, sometimes, never/rarely). We analyzed the items separately for a nuanced picture of physical exercise.

Study 3: Sport Psychological Questionnaires

In Study 3, participants completed nine questionnaires in addition to the VoPE scale (see Table 3). These questionnaires were selected to investigate how the VoPE scale relates to constructs often discussed in the sport psychology literature, shedding light on its nomological network, its validity, and its contribution beyond existing instruments. First, we administered the Behavioral Regulation of Sport Questionnaire (BRSQ; Lonsdale et al., 2008) and the Autotelic Personality Questionnaire (APQ; Tse et al., 2020) to reveal associations between VoPE and different motivation constructs (intrinsic, extrinsic, amotivation, flow). Second, we used the Goal Content for Exercise Questionnaire (GCEQ; Sebire et al., 2008), the Task and Ego Orientation in Sports Questionnaire (TEOSQ; Duda & Nicholls, 1992), and the 2x2 Achievement Goal Questionnaire Sport (AGQS; Conroy et al., 2003) to shed light on potential links between VoPE scores and goals oriented towards learning, improving, and mastering sport performance versus goals oriented towards evaluating, comparing, and avoiding performance. Third, we included the Preference for and Tolerance of the Intensity of Exercise Questionnaire (PRETIEQ; Ekkekakis et al., 2005), the Exercise Addiction Inventory (EAI; Terry et al., 2004), and the Bored of Sports Scale (BOSS; Wolff, Bieleke, et al., 2021). These scales tap into potential relations between the VoPE scale and inclinations to engage in intensive exercise and experience sports-related boredom. Finally, we administered the Mental Toughness Questionnaire (MTQ-10; Papageorgiou et al., 2018) to explore how the VoPE scale relates to performance under challenging circumstances. Taken together, these questionnaires cover a broad range of established constructs related to sports and exercise behavior, allowing us to characterize the VoPE

scale. All questionnaires were presented in random order to avoid order effects.

Behavioral Regulation of Sport Questionnaire (BRSQ). The BRSQ (Lonsdale et al., 2008) is an instrument to assess motivation in a sports context along the lines of self-determination theory. It has nine 4-item scales in which each item starts with "I participate in my sport…": *amotivation* (e.g., "…but I question why I continue"), *external regulation* (e.g., "…because people push me to play"), *introjected regulation* (e.g., "…because I would feel guilty if I quit"), *identified regulation* (e.g., "…because the benefits of sport are important to me"), *integrated regulation* (e.g., "…because it's a part of who I am "), *general intrinsic motivation* (e.g., "…because I enjoy it"), *intrinsic motivation for knowledge* (e.g., "…because I enjoy learning something new about my sport"), *intrinsic motivation for stimulation* (e.g., "…because of the positive feelings that I experience while playing my sport"), and *intrinsic motivation for accomplishment* (e.g., "…because I enjoy the feeling of achievement when trying to reach long-term goals"). Responses are given on a 1 (not at all true) to 5 (very true) Likert scale. The range of internal consistencies was .82 ⩽ ω ⩾ .95.

Autotelic Personality Questionnaire (APQ). The APQ (Tse et al., 2020) is an instrument for assessing attributes of an autotelic personality with 26 items (e.g., "I think the process of completing a task is its own reward"). Responses are given on a 1 (strongly disagree) to 7 (strongly agree) Likert scale. The internal consistency was ω = .88.

Goal Content for Exercise Questionnaire (GCEQ). The GCEQ (Sebire et al., 2008) comprises five 4-item scales measuring the content of goals in exercise settings: *social affiliation* (e.g., "to form close bonds with others"), *image* (e.g., "to improve my appearance"), *health management* (e.g., "to improve my overall health"), *social recognition* (e.g., "to be well thought of by others"), and *skill development* (e.g., "to acquire new exercise skills"). Responses are given on a 1 (not at all important) to 7 (extremely important) Likert scale. The range of internal consistencies was .83 ⩽ ω ⩾ .93.

Task and Ego Orientation in Sports Questionnaire (TEOSQ). The TEOSQ (Duda & Nicholls, 1992) comprises two scales that assess *ego orientation* with 6 items (e.g., "I feel most successful in a sport when others mess up but I do not") and *task orientation* with 7 items (e.g., "I feel most successful in a sport when I do my very best"). Responses are given on a 1 (strongly disagree) to 5 (strongly agree) Likert scale. The range of internal consistencies was .89 ⩽ ω ⩾ .90.

2x2 Achievement Goal Questionnaire Sport (AGQS). The AGQS (Conroy et al., 2003) is a 12-item instrument with four scales measuring *mastery approach* (e.g., "It is important to me to perform as well as I possibly can") and *mastery avoidance* (e.g., "Sometimes I'm afraid that I may not perform as well as I'd like") as well as *performance approach* (e.g., "It is important to me to do well compared to others") and *performance avoidance* goals (e.g., "I just want to avoid performing worse than others"). Responses are given on a 1 (not at all like me) to 7 (completely like me) Likert scale. The range of internal consistencies was .82 ⩽ ω ⩾ .93.

Preference for and Tolerance of the Intensity of Exercise Questionnaire (PRETIEQ). The PRETIEQ (Ekkekakis et al., 2005) is an instrument for assessing the *preference for intensive exercise* (e.g., "The faster and harder the workout, the more pleasant I feel") and *tolerance of intensive exercise* (e.g., "While exercising, I try to keep going even after I feel exhausted") with 8 items per scale. Responses are given on a 1 (I totally disagree) to 5 (I totally agree) Likert scale. The range of internal consistencies was .87 ⩽ ω ⩾ .91.

Exercise Addiction Inventory (EAI). The EAI (Terry et al., 2004) is an instrument for assessing exercise addiction with 6 items (e.g., "Exercise is the most important thing in my life"). Responses are given on a 1 (strongly disagree) to 7 (strongly agree) Likert scale. The internal consistency was ω = .76.

Bored of Sports Scale (BOSS). The BOSS (Wolff, Bieleke, et al., 2021) measures individual differences in trait exercise boredom with 11 items (e.g., "Exercising bores me"). Responses are given on a 1 (strongly disagree) to 5 (strongly agree) Likert scale. The internal consistency was ω = .96.

Mental Toughness Questionnaire (MTQ-10). The MTQ-10 (Papageorgiou et al., 2018) measures individual differences in mental toughness with 10 items ("e.g., I generally cope well with any problems that occur"). Responses are given on a 1 (strongly disagree) to 5 (strongly agree) Likert scale. The range of internal consistency was ω = .91.

Analytic Approach

The analyses were conducted with R (Version 4.2.2; R Core Team, 2022).

Studies 1-3: Basic Psychometric Properties

We hypothesized that the VoPE scale constitutes a unidimensional construct. To examine the one-factor structure of the scale, we conducted confirmatory factor meta-analysis using the *metaSEM* package (Version 1.3.0; Cheung, 2015). We used the two-stage structural equation modeling approach (TSSEM): First, we pooled the correlation matrices obtained in the three studies (total $N = 1364$) based on a fixed-effects model, as we expected that the studies shared the same true effect size. Second, we fitted a one-factorial model with weighted least squares based on the pooled matrix. Additional measurement invariance tests explored construct comparability across the three studies. Model fit was assessed using the root mean square error of approximation (RMSE), the standardized root mean square residual (SRMR), the comparative fit index (CFI) and the Tucker-Lewis index (TLI). Good model fit was assumed for RMSEA ⩽ 0.06, SRMR ⩽ 0.08, and CFI/TLI ⩾ 0.95 (Hu & Bentler, 1999). Invariance was evaluated based on changes in $\Delta\chi^2$ (significance), fit indices (ΔCFI ⩽ 0.01, ΔRMSEA ⩽ 0.015; F. F. Chen, 2007), and fit-complexity tradeoff (ΔBIC; van de Schoot et al., 2012).

To investigate the validity of the VoPE scale in our datasets, we meta-analyzed its associations with external

variables using the *meta* package (Version 6.2-1; Balduzzi et al., 2019) based on a fixed-effects model. Specifically, we tested the correlation (Pearson's *r*) of the VoPE score with exercise volume and the standardized mean difference (Hedges' *g*) in VoPE scores between participants with no competition experience versus some experience.

Study 2: Test-Retest Reliability and Forecasting Behavior

We investigated the test-retest reliability of the VoPE scale by administering it twice in Study 2 (N = 197) with an interval of 5 to 6 weeks between the two assessments. We then conducted a Bland-Altman analysis (Bland & Altman, 1999), in which the individual means of the two assessments are plotted against the differences between them. This yields insight into potential bias (i.e., non-zero difference between the measures) and how it changes as a function of the mean of the measures. Here, we show the Bland-Altman plot and report the bias of the test and retest scores of the VoPE scale along with its 95% confidence interval. We additionally determined the intraclass correlation coefficient (ICC) with the ICC() function of the *psych* package (Version 2.3.3, Revelle, 2023) using a single-rating, absolute-agreement, two-way mixed-effect model (Koo & Li, 2016). An ICC of less than .50 indicates poor test-retest reliability, ICCs between .50 and .75 indicate moderate reliability, ICCs between .75 and .90 indicate good reliability, and ICCs greater than .90 indicate excellent reliability (Koo & Li, 2016).

To examine the predictive validity of the VoPE scale in Study 2, we regressed measures of physical (in-)activity assessed with the IPAQ and exercise volume measured with GSLTPAQ assessed at T2 on the VoPE scale assessed at T1. We report standardized regression coefficients along with their standard errors and the 95% confidence intervals, as well as a robustness check in which we adjusted for sociodemographic variables.

Study 3: Psychometric Network Modeling and Out-of-Sample Predictions

We relied on a psychometric network approach (for an overview, see Borsboom et al., 2021) to explore the nomological network of the VoPE scale in Study 3 (N = 805). Psychometric networks display the strength of conditional associations (edges) between a set of variables (nodes). We used the EBICglasso() and the qgraph() function of the *qgraph* package (Version 1.9.3; Epskamp et al., 2012) to estimate and visualize the network. Estimation relies on the graphical least absolute shrinkage and selection operator (GLASSO; Friedman et al., 2008) to estimate regularized partial correlation networks. It then selects a model based on the extended Bayesian Information Criterion (EBIC; J. Chen & Chen, 2008), for which we set the hyperparameter γ to 0.5. To examine the importance of the nodes in the network, we determined different centrality indices (see Epskamp et al., 2018): strength (i.e., the magnitude of the direct connections of a node to other nodes), closeness (i.e., how quickly a node can reach all other nodes in the network), betweenness (i.e., how often a node lies on the shortest path between two other nodes). We additionally report the expected influence (i.e., the influence of a node on other nodes in the networks), which is recommended when a network contains both positive and negative edges (Robinaugh et al., 2016). Finally, we used the bootnet() function of the *bootnet* package (Version 1.5; Epskamp et al., 2018) to bootstrap the network 1,000 times to determine the 95% confidence intervals around edge weights involving the VoPE scale.

Next, we investigated which scale best predicts participants' experience with sports competitions in our data (binary coded as 0 = no experience and 1 = some experience). We chose experience with sports competitions as the criterion for several reasons. First, participating in competitions requires a level of performance that is difficult to reconcile with a purely instrumental view on effort (e.g., training to impress others, being active to stay healthy), making it plausible that the value of effort plays a role. Second, it should be straightforward for people to recall and report their competition experience, increasing confidence in the validity of the criterion. Third, experience with competitions has been used as a criterion in the development and validation of other scales (e.g., Schüler et al., 2023), facilitating a comparison with findings in the present research.

We turned to a machine learning approach, which is becoming increasingly popular in psychological science when the focus is on out-of-sample predictions and generalizability (for an overview, see Pargent et al., 2023; Yarkoni & Westfall, 2017). Specifically, we used regularized regression, which penalizes the number of coefficients in a regression model by effectively shrinking them towards zero (ridge), forcing them to zero (lasso) or a mixture of both (elastic net; Hastie et al., 2009). Regularized regression is implemented in the package *caret* (Version 6.0-93; Kuhn, 2022). Here, we specified participants' standardized scores on each scale as the feature matrix of the regression model. We then randomly split the data in a training set (80%) and a test set (20%). A Generalized Linear Model (GLM) was trained with 10-fold leave-one-out cross-validation (LOO-CV; Bischl et al., 2012). We performed hyperparameter tuning with a grid search for the hyperparameters α (between 0 and 1) and λ (between 10^{-5} and 10^{5}). We report the confusion matrix for the best performing model along with several performance metrics (e.g., accuracy, sensitivity, specificity) to quantify the model's performance in terms of out-of-sample prediction. We then determined the importance of the questionnaires and scales in the model using the (normalized) absolute values of the model coefficients.

Results

Studies 1-3: Psychometric Properties of the VoPE Scale

Figure 1a shows the distribution of VoPE scores across studies (N = 1364). There are two noteworthy observations: First, participants across all three studies tended to value physical effort on average, as indicated by a moderate agreement with the items on average (pooled M = 4.66, SD

= 1.42). Second, we observed a larger number of low VoPE scores in Study 2, which drew upon a more representative sample compared to a student sample (Study 1) and a sample screened for a minimum level of activity (Study 3). This suggests that the VoPE scale is sensitive for the entire spectrum from not valuing physical effort at all to valuing it highly.

The internal consistency of the scale was good (pooled Ω = 0.89, 95% CI [0.88, 0.90]). A confirmatory factor meta-analysis (meta-CFA; Figure 1b) suggested that all four items loaded on a common factor across studies, $\chi^2(2) = 0.54$, $p = 0.763$, CFI = 1, TLI = 1, SRMR = 0.003, RMSEA = 0.000, 95% CI [0.000, 0.036], attesting to the unidimensional structure of the VoPE scale. The standardized factor loadings ranged from 0.66 to 0.91, indicating that all four items substantially contribute to the VoPE score.

Configural, metric, and scalar invariance models fit well (CFI ⩾ 0.990, TLI ⩾ 0.982, RMSEA ⩽ 0.064, SRMR ⩽ 0.038). Although $\Delta\chi^2$ was significant (metric: $\Delta\chi^2(6) = 27.11$, $p < .001$; scalar: $\Delta\chi^2(6) = 19.26$, $p = .004$), changes in fit indices were mostly within acceptable limits (metric: ΔCFI = -0.006, ΔRMSEA = 0.060; scalar: ΔCFI = -0.004, ΔRMSEA = 0.004) and the tradeoff between model fit and model complexity consistently improved (metric: ΔBIC = -15.74, scalar: ΔBIC = -23.38), indicating construct comparability across studies.

Attesting to the validity of the scale in our data, higher VoPE scores were associated with higher weekly exercise volume (frequency x duration; Figure 1c). In line with the correlations in each study, the pooled correlation was positive and significant (Figure 1e), $r = .32$, 95% CI [.27, .37], $z = 12.31$, $p < .001$. Additionally, we observed higher VoPE sores among people with some sports competition experience compared to participants with no experience (Figure 1d). In line with the standardized mean difference in each study, the pooled standardized mean difference was significant (Figure 1f), SMD = 0.63, 95% CI [0.49, 0.78], $z = 8.68$, $p < .001$. Finally, we observed significant pooled correlations of VoPE score with age, $r = -0.09$, 95% CI [−.14, −.04], $z = 3.41$, $p < .001$, gender (0 = female, 1 = male), $r = 0.22$, 95% CI [.17, .27], $z = 8.19$, $p < .001$, income, $r = 0.19$, 95% CI [.13, .25], $z = 6.41$, $p < .001$, education, $r = 0.09$, 95% CI [.03, .14], $z = 12.31$, $p = .005$, and subjective SES, $r = 0.19$, 95% CI [.13, .24], $z = 6.21$, $p < .001$. Most individual correlations were in line with these meta-analytic findings, except non-significant links between VoPE and age in Study 1 and between VoPE and education in Study 2 (see OSF for detailed results).

Study 2: Test-Retest Reliability and Forecasting

We examined the test-retest reliability and the predictive validity of the VoPE scale in Study 2 ($N = 197$). We found a moderate correlation between the VoPE scores over the course of 5 to 6 weeks, ICC = 0.77, 95% CI [.71, .82], attesting to the test-retest reliability of the VoPE scale. The Bland-Altman-Plot (Figure 2) shows that the difference in VoPE scores between the two measurements was normally distributed around a non-significant bias of 0.004, 95% CI [−0.149, 0.157] across the possible values of the VoPE scale.

Next, we turned to forecasting and regressed measures of physical activity from the IPAQ (vigorous activity, moderate activity, walking, sitting) and measures of physical exercise from the GSLTPAQ (strenuous exercise, moderate exercise, mild exercise, and sweating) on the VoPE score measured several weeks before. As shown in Table 4, the VoPE score predicted more vigorous and more moderate physical activity as well as more walking 5 to 6 weeks later, and less time spent sitting. Moreover, the VoPE score predicted strenuous, moderate, and mild physical exercise several weeks later as well as more occasions of working up a sweat through exercise. In line with the conceptualization of the value of physical effort, the VoPE score explained more variance for vigorous forms of activity (15%) and exercise (19%) as well as occasions of working up a sweat (16%) than of moderate or mild forms of activity (9%) and exercise (2 – 11%). Adjusting for sociodemographic variables (age, gender, income, education, employment, and subjective socio-economic status) had negligible effects on our results. The only change in terms of significance was that the VoPE score no longer predicted mild physical exercise measured with the GSLTPAQ, β = 0.13, SE = 0.09, p = 0.138, 95% CI [−0.04; 0.30].

Study 3: Psychometric Network and Out-of-Sample Prediction

The estimated psychometric network consisted of 27 nodes representing the questionnaires and their scales (Figure 3a). It comprised 160 non-zero edges out of 351 (45.6%) with a mean weight of 0.026. Higher scores on the VoPE scale were associated especially with a higher preference for exercise intensity (PRETIEQ-PREF), a higher inclination for exercise addiction (EAI), and a higher intrinsic motivation for stimulation (BRSQ-IMST), and less exercise-related boredom (BOSS). The magnitude of the regularized partial correlations did not exceed values of about .2 (Figure 3b), indicating that the VoPE scale makes unique contributions over and above the set of sport psychological questionnaires used in the present research. Figure 3c additionally suggests that the VoPE scale has a rather pronounced betweenness centrality (i.e., it is often located on the shortest path between two other nodes), while other centrality measures had average levels. To facilitate comparisons with the existing literature, we additionally determined bivariate correlations between all scales and provide them on OSF.

The regularization with the elastic net regression shrunk 12 of 27 coefficients to zero (44%). Of the non-zero coefficients in our data (see Figure 3d), the VoPE scale emerged as the most important predictor of competition experience ($b = 0.29$, OR = 1.34), followed by tolerance for the intensity of effort ($b = 0.18$, OR = 1.20) and performance approach goals ($b = 0.18$, OR = 1.19). Other important predictors of competition experience were general intrinsic motivation ($b = 0.10$, OR = 1.11) as well as intrinsic motivation for stimulation ($b = 0.11$, OR = 1.12) and knowledge ($b = 0.07$ OR = 1.07), and social affiliation goals ($b = 0.08$, OR = 1.08). Introjected regulation ($b = -0.11$, OR = 0.90), task orientation ($b = -0.09$, OR = 0.91), performance avoidance orientation ($b = -0.05$, OR = 0.95), and exercise-related boredom ($b =$

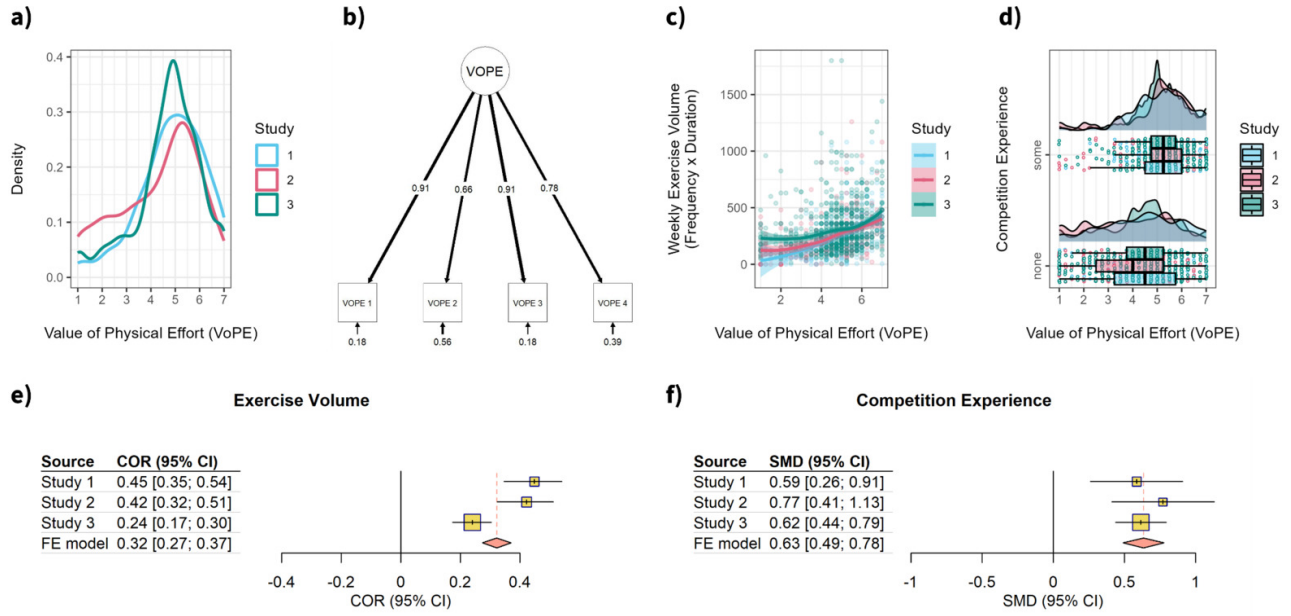

Figure 1. Establishing the basic psychometric properties of the VoPE Scale across Studies 1, 2, and 3.

Note. **a)** Density plot showing the distribution of VoPE scores in Studies 1 to 3. **b)** Results of the confirmatory meta-analysis with standardized factor loadings. **c)** Scatterplot of the weekly exercise volume as a function of the VoPE score. Lines represent locally estimated scatterplot smoothing (LOESS). **d)** Raincloud plots showing the distribution of VoPE scores for participants with no versus some competition experience in Studies 1 to 3. **e), f)** Forest plots of the individual (squares) and the pooled (diamonds) association between VoPE scores and exercise volume (Pearson's *r*) and competition experience (Hedges *g*), respectively. Total $N = 1364$.

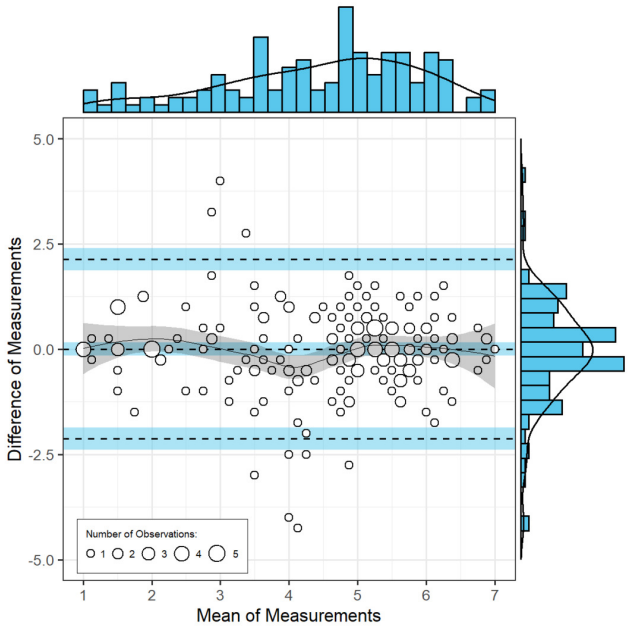

Figure 2. Bland-Altman plot showing correspondence between the test-retest scores of the VoPE scale over the course of 5 to 6 weeks in Study 2.

Note. The shaded regions correspond to 95% confidence intervals. N = 197.

−0.05, OR = 0.95) predicted a lower likelihood of competition experience.

Turning to the out-of-sample prediction, the most accurate model was associated with the hyperparameters $\alpha = 0.2$ and $\lambda = 0.056$. The model achieved an accuracy of 0.62, 95% CI [0.54, 0.70], which was significantly higher than the no-information rate of 0.55, $p = .047$ (Figure 3e). The model had a sensitivity of 47.2% and a specificity of 74.2%.

Discussion

Across three independent studies, we have developed and validated the Value of Physical Effort (VoPE) scale. The instrument demonstrated desirable psychometric properties in terms of reliability and validity in all datasets. Moreover, the VoPE scale performed well in forecasting self-reported sport and exercise behavior and made the most substantial contribution to a model for out-of-sample predictions regarding experience in sporting competitions. The instrument allows researchers to assess individual differences ranging from not valuing physical effort at all to valuing it highly.

Implications

Our research fills an important gap in the still young literature on the value of effort, which has traditionally focused on the cognitive domain (e.g., Inzlicht et al., 2018). This work has been inspired by the need for cognition scale (Cacioppo & Petty, 1982), which allows researchers to assess individual differences in the value of cognitive effort. In contrast, relatively little attention has been paid to the value of effort in the physical domain (Stähler et al., 2025), which is likely due to a lack of measurement instruments. By introducing the VoPE scale, we demonstrated that people differ in their valuation of physical effort as well, and that these individual differences are related to sports behavior and outcomes, exercise volume, and physical activity. Our findings are consistent with research using the

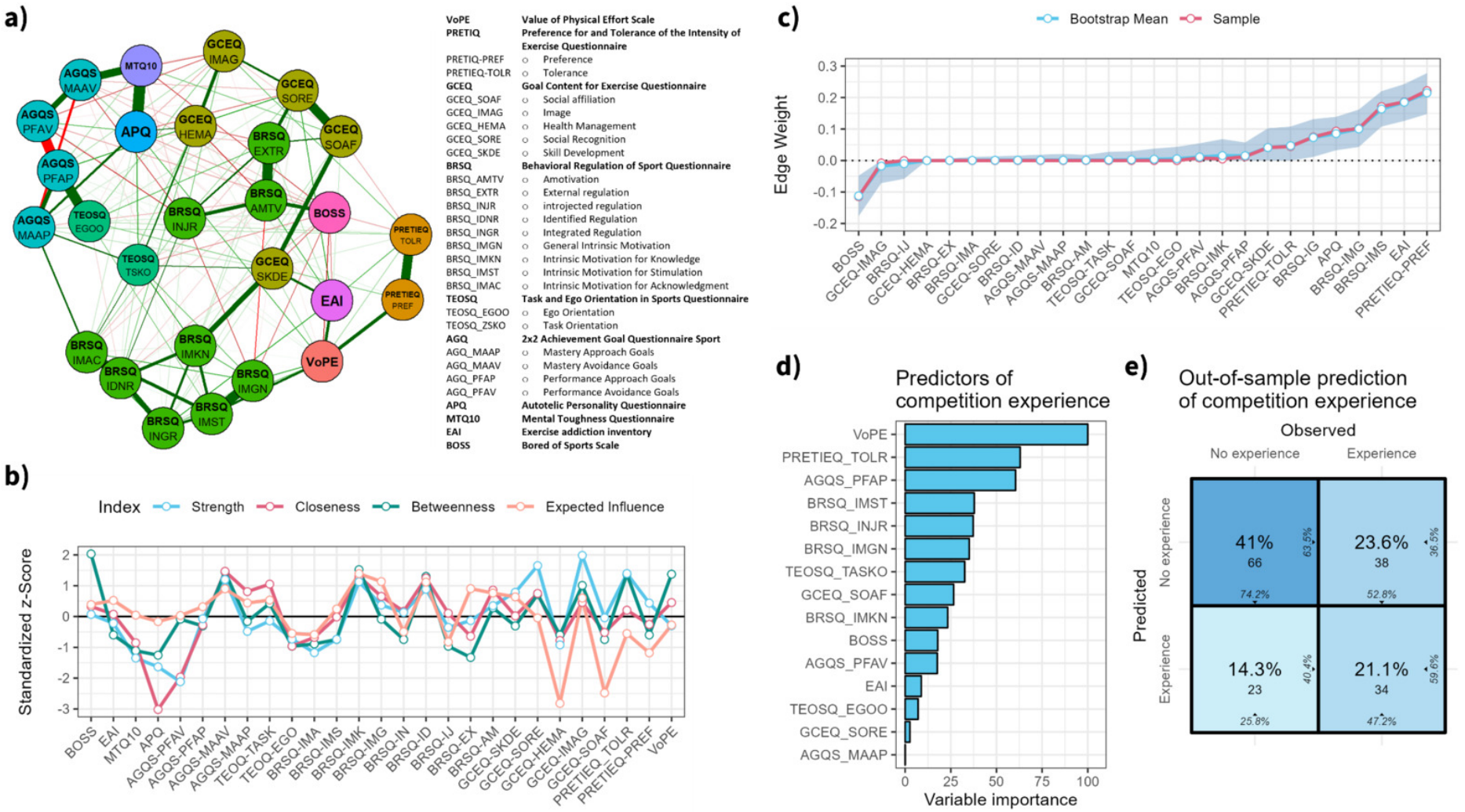

Figure 3. Characterizing the psychometric network of the VoPE scale and examining its performance in out-of-sample predictions in Study 3.

Note. **a)** Psychometric network of the questionnaires and scales included in Study 3. The layout is based on the Fruchterman-Reingold algorithm implemented in the *qgraph* package. Nodes represent questionnaires and scales, edges represent regularized partial correlations. Line thickness and color code the magnitude (thicker = stronger) and direction (green = positive, red = negative) of the association. **b)** Standardized z-scores of the four centrality indices determined for each node in the psychometric network. **c)** Averaged bootstrapped and sample values of the weights of the edges connecting the VoPE scale with all other questionnaires and scales. The shaded area represents the bootstrapped 95% interval. **d)** Variable importance in terms of the normalized absolute model coefficients of the best performing model predicting competition experience. **e)** Confusion matrix for the best performing model in the out-of-sample prediction of competition experience. The arrows in each panel indicate the row-wise or column-wise computation of percentages. N = 805.

Table 4. Forecasting physical activity and exercise behavior using VoPE scores in Study 2.

	Physical Activity (IPAQ) at T2				Physical Exercise (GSLTPAQ) at T2			
	Vigorous	Moderate	Walking	Sitting	Strenuous	Moderate	Mild	Sweating
β	0.39	0.30	0.30	−0.25	0.44	0.33	0.14	0.40
SE	0.07	0.07	0.07	0.07	0.06	0.07	0.07	0.07
95% CI								
LL	0.26	0.17	0.16	−0.39	0.31	0.20	0.00	0.27
UL	0.52	0.44	0.43	−0.12	0.56	0.46	0.28	0.53
p	<.001	<.001	<.001	<.001	<.001	<.001	.047	<.001
R^2	0.15	0.09	0.09	0.06	0.19	0.11	0.02	0.16
adj. R^2	0.15	0.09	0.08	0.06	0.19	0.10	0.02	0.16
N	194	194	194	197	197	197	197	197

Note. β = standardized coefficient, SE = standard error, CI = confidence interval, LL = lower limit, UL = upper limit.

recently introduced Physical Effort Scale (PES; Cheval et al., 2024), which demonstrates individual differences in approach and avoidance tendencies towards physical effort. The development of instruments tapping into various facets of liking effort leverages research on the commonalities and differences between the value of cognitive versus physical effort (Wolff et al., 2024).

We used a psychometric network approach to explore empirically how the value of physical effort relates to constructs commonly used in sport psychological research. The strongest associations emerged with the preference for exercise intensity (Ekkekakis et al., 2005) and the inclination for exercise dependence (Terry et al., 2004). We also found evidence for a negative link between the value of physical effort and exercise-related boredom (Wolff, Bieleke, et al., 2021). Boredom is a value-based experience (Martarelli et al., 2023), and most theories of boredom assume that low levels of value give rise to boredom (for an overview, see Bieleke et al., 2024). This implies that people should get bored by effortful sports and exercise activities when they ascribe little value to the physical effort that is commonly involved in these activities. Further, people tend to avoid and escape boring activities (Bieleke et al., 2022), making boredom a likely mechanism involved in the negative association between VoPE scores and sports and exercise behavior.

The longitudinal design of Study 2 allowed us to gauge the stability of the VoPE scale over time. The observed correlation was moderate, suggesting that to some extent the VoPE scale captures changes in the value of physical effort. Indeed, it is plausible to assume that value changes: For instance, the learned industriousness theory asserts that effort becomes valuable when it is repeatedly followed by rewards (Eisenberger, 1992). Conversely, it should become less valuable when no reward is available. It would be interesting to test whether the VoPE scale is sensitive to changes in value, for instance, in experimental manipulations of effort-contingent rewards. On a related note, accounting for individual or contextual differences in the value of physical effort might help to advance theoretical models of physical activity behavior, such as the theory of effort minimization in physical activity (Cheval & Boisgontier, 2021). In line with predictions derived from this theory, a stronger inclination to approach (vs. avoid) physical effort has already been shown to foster the link between the intention to be physically active and physical activity (Maltagliati et al., 2024). It is plausible that people who assign greater value to physical effort also tend to approach effort more readily. Testing ideas like this and locating the value of physical effort in theoretical models is thus a promising endeavor for future research.

Finally, our results suggest that viewing effort solely as costly and explaining its expenditure in purely instrumental terms might be an overly narrow perspective. Our results suggest that some people value the physical effort involved in sports and exercise to some degree, possibly because effort is commonly followed by rewards, making it a secondary reinforcer (Eisenberger, 1992). This is in line with research showing that consistently rewarding effort can lead people to choose more effortful activities (e.g., Clay et al., 2022), although valuing effort for its own sake still remains the exception rather than the rule (e.g., David et al., 2022; Wolff et al., 2024). What contributes to the costs and value people associate with effort thus remains an open question, and the VoPE scale hopefully assists researchers in disentangling these contributions further.

Limitations and Future Directions

First, the VoPE scale is a self-report instrument and should therefore be complemented by objective indicators of behavior in future research. For instance, one might use behavioral tasks in which participants make enforceable decisions to invest physical effort and relate these decisions to their answers in the VoPE scale. This would be a promising step for future research, as linking VoPE scores to behavior in standardized tasks, such as the Effort Expenditure for Rewards Task (Treadway et al., 2009), the Progressive Ratio Task (Hodos, 1961), or the Ring Task for effort preferences (Wolff et al., 2024) would further advance our understanding of how a dispositional valuation of physical effort exerts its effects on human behavior. In addition, VoPE scores might be correlated with indicators of sports and exercise behavior or physical activity that are measured by devices such as sports and fitness watches. For instance,

collecting accelerometer data and conducting ecological momentary assessments could provide valuable insights into the weak link between VoPE scores and self-reported mild forms of exercise and physical activity observed in this study.

Second, we used a rather broad framing of physical effort in terms of sports for designing the VoPE scale in the present research. However, it is plausible that people value different aspects of physical effort differently. For instance, runners might ascribe particularly high value to the effort involved in endurance performance, while weightlifters might ascribe higher value to the effort involved in strength performance. Also, we found that the VoPE scale explained little variance of mild forms of exercise and activities, which might be due to its focus on sports rather than on physical activity. Future research might adapt the VoPE scale to different forms of physical effort to scrutinize potential inter- and intraindividual differences in their valuations.

Third, the present research capitalized on student and adult samples without dedicated sports experience. However, it is well-known that elite athletes differ in their physical (e.g., maximal power output; Lorenz et al., 2013) and mental configuration (e.g., trait self-control; Wolff et al., 2019) from non-elite athletes, and these differences are likely to be even more pronounced in comparison to hobby athletes. Therefore, it seems promising to compare athletes at different performance levels with regard to their answers on the VoPE scale. Similarly, investigating changes in the VoPE scale in these populations over longer time scales would be worthwhile.

Conclusion

The VoPE scale is a short and economical instrument for assessing individual differences in the value of physical effort. It can be administered to investigate differences in the valuation of various kinds of effort (e.g., cognitive versus physical) and our data suggest it predicts behavior involving physical effort (e.g., sports and exercise, physical activity). Moreover, the VoPE scale could be used to measure changes in the value of physical effort and determine the effectiveness of interventions in future research. Finally, the VoPE scale help to shed light on various value-based constructs, such as exercise-related boredom, and their consequences for behavior.

Author Contributions

Maik Bieleke: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing–Original Draft, Writing–Review & Editing, Funding Acquisition; **Johanna Stähler:** Conceptualization, Methodology, Writing–Review & Editing; **Wanja Wolff:** Conceptualization, Methodology, Writing–Original Draft, Writing–Review & Editing; **Julia Schüler:** Conceptualization, Methodology, Investigation, Writing–Review & Editing, Funding Acquisition

Data Availability

Materials, data, and scripts used in this manuscript are available at OSF (https://osf.io/zk9qr/).

Funding

The authors gratefully acknowledge funding from the Committee on Research (AFF) at the University of Konstanz for the research initiative "ProPELL: Promoting Physical Exercise in Lab and Life". Maik Bieleke was supported by an Independent Research Grant awarded by the Institute for Advanced Study for Early Career Researchers at the University of Konstanz.

Competing Interests

The authors declare no competing interests.

Ethics Statement

The study protocol and measurements were approved by the Ethics Committee at the University of Konstanz (approval IRB25KN003-04/w).

Editors: Eunike Wetzel (Senior Editor)

Submitted: February 25, 2025 PDT. Accepted: May 10, 2025 PDT. Published: June 30, 2025 PDT.

References

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence Based Mental Health*, *22*(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Bieleke, M., Ripper, L., Schüler, J., & Wolff, W. (2022). Boredom is the root of all evil—or is it? A psychometric network approach to individual differences in behavioural responses to boredom. *Royal Society Open Science*, *9*(9), 211998. https://doi.org/10.1098/rsos.211998

Bieleke, M., Wolff, W., & Martarelli, C. S. (Eds.). (2024). *Handbook of boredom*. Routledge.

Bischl, B., Mersmann, O., Trautmann, H., & Weihs, C. (2012). Resampling methods for meta-model validation with recommendations for evolutionary computation. *Evolutionary Computation*, *20*(2), 249–275. https://doi.org/10.1162/EVCO_a_00069

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Borsboom, D., Deserno, M. K., Rhemtulla, M., Epskamp, S., Fried, E. I., McNally, R. J., Robinaugh, D. J., Perugini, M., Dalege, J., Costantini, G., Isvoranu, A.-M., Wysocki, A. C., Van Borkulo, C. D., Van Bork, R., & Waldorp, L. J. (2021). Network analysis of multivariate data in psychological science. *Nature Reviews Methods Primers*, *1*(1), 58. https://doi.org/10.1038/s43586-021-00055-w

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*(1), 109–131. https://doi.org/10.1146/annurev.ps.40.020189.000545

Brinkmann, K., Richter, M., & Gendolla, G. H. E. (2021). The intensity side of volition: A theoretical and empirical overview of effortful striving. *Zeitschrift für Sportpsychologie*, *28*(3), 97–108. https://doi.org/10.1026/1612-5010/a000323

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Carbonaro, W. (2005). Tracking, students' effort, and academic achievement. *Sociology of Education*, *78*(1), 27–49. https://doi.org/10.1177/003804070507800102

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, J., & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, *95*(3), 759–771. https://doi.org/10.1093/biomet/asn034

Cheung, M. W.-L. (2015). metaSEM: An R package for meta-analysis using structural equation modeling. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01521

Cheval, B., & Boisgontier, M. P. (2021). The theory of effort minimization in physical activity. *Exercise and Sport Sciences Reviews*, *49*(3), 168–178. https://doi.org/10.1249/JES.0000000000000252

Cheval, B., Maltagliati, S., Courvoisier, D. S., Marcora, S., & Boisgontier, M. P. (2024). Development and validation of the Physical Effort Scale (PES). *Psychology of Sport and Exercise*, *72*, 102607. https://doi.org/10.1016/j.psychsport.2024.102607

Clay, G., Mlynski, C., Korb, F. M., Goschke, T., & Job, V. (2022). Rewarding cognitive effort increases the intrinsic value of mental labor. *Proceedings of the National Academy of Sciences*, *119*(5), e2111785119. https://doi.org/10.1073/pnas.2111785119

Conroy, D. E., Elliot, A. J., & Hofer, S. M. (2003). A 2 × 2 achievement goals questionnaire for sport: Evidence for factorial invariance, temporal stability, and external validity. *Journal of Sport and Exercise Psychology*, *25*(4), 456–476. https://doi.org/10.1123/jsep.25.4.456

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International Physical Activity Questionnaire: 12-country reliability and validity. *Medicine & Science in Sports & Exercise*, *35*(8), 1381–1395. https://doi.org/10.1249/01.MSS.0000078924.61453.FB

David, L., Vassena, E., & Bijleveld, E. (2022). The aversiveness of mental effort: A meta-analytic review of the association between mental effort and negative affect. *PsyArXiv*. Preprint. https://doi.org/10.31234/osf.io/m8zf6

Duda, J. L., & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology*, *84*(3), 290–299. https://doi.org/10.1037/0022-0663.84.3.290

Eisenberger, R. (1992). Learned industriousness. *Psychological Review*, *99*(2), 248–267. https://doi.org/10.1037/0033-295X.99.2.248

Ekkekakis, P., Hall, E. E., & Petruzzello, S. J. (2005). Some like it vigorous: Measuring individual differences in the preference for and tolerance of exercise intensity. *Journal of Sport and Exercise Psychology*, *27*(3), 350–374. https://doi.org/10.1123/jsep.27.3.350

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., Cramer, A. O. J., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4). https://doi.org/10.18637/jss.v048.i04

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Godin, G., & Shephard, R. J. (1985). A simple method to assess exercise behavior in the community. *Canadian Journal of Applied Sport Sciences. Journal Canadien Des Sciences Appliquees Au Sport*, *10*(3), 141–146.

Halperin, I., & Vigotsky, A. D. (2023). A conceptual framework of effort and perception of effort. *PsyArXiv*. Preprint. https://doi.org/10.31234/osf.io/h4dfe

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Haynos, A. F., Koithan, E., & Hagan, K. E. (2022). Learned industriousness as a translational mechanism in anorexia nervosa. *Nature Reviews Psychology*, *2*(2), 112–126. https://doi.org/10.1038/s44159-022-00134-z

Hodos, W. (1961). Progressive ratio as a measure of reward strength. *Science*, *134*(3483), 943–944. https://doi.org/10.1126/science.134.3483.943

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, *22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*, *143*(1), 131–141. https://doi.org/10.1037/a0031048

Kuhn, M. (2022). *caret: Classification and regression training*. https://CRAN.R-project.org/package=caret

Kurzban, R. (2016). The sense of effort. *Current Opinion in Psychology*, *7*, 67–70. https://doi.org/10.1016/j.copsyc.2015.08.003

Lindwall, M., Barkoukis, V., Grano, C., Lucidi, F., Raudsepp, L., Liukkonen, J., & Thøgersen-Ntoumani, C. (2012). Method effects: The problem with negatively versus positively keyed items. *Journal of Personality Assessment*, *94*(2), 196–204. https://doi.org/10.1080/00223891.2011.645936

Lonsdale, C., Hodge, K., & Rose, E. A. (2008). The Behavioral Regulation in Sport Questionnaire (BRSQ): Instrument development and initial validity evidence. *Journal of Sport and Exercise Psychology*, *30*(3), 323–355. https://doi.org/10.1123/jsep.30.3.323

Lorenz, D. S., Reiman, M. P., Lehecka, B. J., & Naylor, A. (2013). What performance characteristics determine elite versus nonelite athletes in the same sport? *Sports Health: A Multidisciplinary Approach*, *5*(6), 542–547. https://doi.org/10.1177/1941738113479763

Maltagliati, S., Raichlen, D., Rhodes, R., & Cheval, B. (2024). *Closing the intention-behavior gap in physical activity: The moderating effect of individual differences in the valuation of physical effort*. https://doi.org/10.51224/SRXIV.375

Martarelli, C. S., Berthouzoz, P., Bieleke, M., & Wolff, W. (2023). Bored of sports? Investigating the interactive role of engagement and value as predictors of boredom in athletic training. *Sport, Exercise, and Performance Psychology*, *12*(2), 141–154. https://doi.org/10.1037/spy0000320

Maxcy, J., Wicker, P., & Prinz, J. (2019). Happiness as a reward for torture: Is participation in a long-distance triathlon a rational choice? *Journal of Sports Economics*, *20*(2), 177–197. https://doi.org/10.1177/1527002518758144

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., Sears, M. R., Thomson, W. M., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Papageorgiou, K. A., Malanchini, M., Denovan, A., Clough, P. J., Shakeshaft, N., Schofield, K., & Kovas, Y. (2018). Longitudinal associations between narcissism, mental toughness and school achievement. *Personality and Individual Differences*, *131*, 105–110. https://doi.org/10.1016/j.paid.2018.04.024

Pargent, F., Schoedel, R., & Stachl, C. (2023). Best practices in supervised machine learning: A tutorial for psychologists. *Advances in Methods and Practices in Psychological Science*, *6*(3), 25152459231162559. https://doi.org/10.1177/25152459231162559

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revelle, W. (2023). *psych: Procedures for psychological, psychometric, and personality research*. https://CRAN.R-project.org/package=psych

Rewitz, K., Schindler, S., & Wolff, W. (2024). Examining the alignment between subjective effort and objective force production. *PloS One*, *19*(8), Article e0307994. https://doi.org/10.1371/journal.pone.0307994

Rhodes, R. E., & Dickau, L. (2012). Experimental evidence for the intention–behavior relationship in the physical activity domain: A meta-analysis. *Health Psychology*, *31*(6), 724–727. https://doi.org/10.1037/a0027290

Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology*, *125*(6), 747–757. https://doi.org/10.1037/abn0000181

Schüler, J., Stähler, J., & Wolff, W. (2023). Mind-over-body beliefs in sport and exercise: A driving force for training volume and performance, but with risks for exercise addiction. *Psychology of Sport and Exercise*, *68*, 102462. https://doi.org/10.1016/j.psychsport.2023.102462

Sebire, S. J., Standage, M., & Vansteenkiste, M. (2008). Development and validation of the Goal Content for Exercise Questionnaire. *Journal of Sport and Exercise Psychology*, *30*(4), 353–377. https://doi.org/10.1123/jsep.30.4.353

Silm, G., Pedaste, M., & Täht, K. (2020). The relationship between performance and test-taking effort when measured with self-report or time-based instruments: A meta-analytic review. *Educational Research Review*, *31*, 100335. https://doi.org/10.1016/j.edurev.2020.100335

Stähler, J., Bieleke, M., Wolff, W., & Schüler, J. (2025). Different functions of physical effort: A scoping review of the value of physical effort in physical activity and sports. *Motivation and Emotion*, *49*, 259–283. https://doi.org/10.1007/s11031-025-10123-3

Terry, A., Szabo, A., & Griffiths, M. (2004). The Exercise Addiction Inventory: A new brief screening tool. *Addiction Research & Theory*, *12*(5), 489–499. https://doi.org/10.1080/16066350310001637363

Therriault, D. J., Redifer, J. L., Lee, C. S., & Wang, Y. (2015). On cognition, need, and action: How working memory and need for cognition influence leisure activities. *Applied Cognitive Psychology*, *29*(1), 81–90. https://doi.org/10.1002/acp.3078

Treadway, M. T., Buckholtz, J. W., Schwartzman, A. N., Lambert, W. E., & Zald, D. H. (2009). Worth the 'EEfRT'? The Effort Expenditure for Rewards Task as an objective measure of motivation and anhedonia. *PLoS ONE*, *4*(8), e6598. https://doi.org/10.1371/journal.pone.0006598

Tse, D. C. K., Lau, V. W., Perlman, R., & McLaughlin, M. (2020). The development and validation of the Autotelic Personality Questionnaire. *Journal of Personality Assessment*, *102*(1), 88–101. https://doi.org/10.1080/00223891.2018.1491855

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

Van Iddekinge, C. H., Arnold, J. D., Aguinis, H., Lang, J. W. B., & Lievens, F. (2023). Work effort: A conceptual and meta-analytic review. *Journal of Management*, *49*(1), 125–157. https://doi.org/10.1177/01492063221087641

Wang, Y., & Ashokan, K. (2021). Physical exercise: An overview of benefits from psychological level to genetics and beyond. *Frontiers in Physiology*, *12*, 731858. https://doi.org/10.3389/fphys.2021.731858

Wolff, W., Bertrams, A., & Schüler, J. (2019). Trait self-control discriminates between youth football players selected and not selected for the german talent program: A bayesian analysis. *Frontiers in Psychology*, *10*, 2203. https://doi.org/10.3389/fpsyg.2019.02203

Wolff, W., Bieleke, M., Stähler, J., & Schüler, J. (2021). Too bored for sports? Adaptive and less-adaptive latent personality profiles for exercise behavior. *Psychology of Sport and Exercise*, *53*, 101851. https://doi.org/10.1016/j.psychsport.2020.101851

Wolff, W., Hirsch, A., Bieleke, M., & Shenhav, A. (2021). Neuroscientific approaches to self-regulatory control in sports. In C. Englert & I. Taylor (Eds.), *Motivation and self-regulation in sport and exercise* (pp. 149–165). Routledge. https://doi.org/10.4324/9781003176695-11

Wolff, W., Stähler, J., Schüler, J., & Bieleke, M. (2024). On the specifics of valuing effort: A developmental and a formalized perspective on preferences for mental and physical effort. *Peer Community Journal*, *4*, e78. https://doi.org/10.31234/osf.io/ycvxw

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Zhang, X., Noor, R., & Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PLOS ONE*, *11*(6), e0157795. https://doi.org/10.1371/journal.pone.0157795

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press.

Supplementary Materials

Peer Review Communication

Download: https://collabra.scholasticahq.com/article/140736-development-and-validation-of-the-value-of-physical-effort-vope-scale/attachment/288463.docx?auth_token=-P638FI7gN1uT1NlNVvx

Response Letter

Download: https://collabra.scholasticahq.com/article/140736-development-and-validation-of-the-value-of-physical-effort-vope-scale/attachment/288464.pdf?auth_token=-P638FI7gN1uT1NlNVvx