Measuring Motivations of Crowdworkers: The Multidimensional Crowdworker Motivation Scale

LISA POSCH, GESIS–Leibniz Institute for the Social Sciences, Germany and Graz University of Technology, Austria

ARNIM BLEIER and CLEMENS M. LECHNER, GESIS–Leibniz Institute for the Social Sciences, Germany

DANIEL DANNER, University of Applied Labour Studies, Germany

FABIAN FLÖCK, GESIS–Leibniz Institute for the Social Sciences, Germany

MARKUS STROHMAIER, RWTH Aachen University, Germany and GESIS – Leibniz Institute for the Social Sciences, Germany

Crowd employment is a new form of short-term and flexible employment that has emerged during the past decade. To understand this new form of employment, it is crucial to illuminate the underlying motivations of the workforce involved in it. This article introduces the Multidimensional Crowdworker Motivation Scale (MCMS), a scale for measuring the motivation of crowdworkers on microtask platforms. The MCMS is theoretically grounded in self-determination theory and tailored specifically to the context of paid crowdsourced microlabor. The scale measures the motivation of crowdworkers along six motivational dimensions, ranging from amotivation to intrinsic motivation. We validated the MCMS on data collected in ten countries and three income groups. Factor analyses demonstrated that the MCMS's six dimensions showed good model fit, validity, and reliability. Furthermore, our measurement invariance tests showed that motivations measured with the MCMS are comparable across countries and income groups, and we present a first cross-country comparison of crowdworker motivations. This work constitutes an important first step toward understanding the motivations of the international crowd workforce.

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**; • **General and reference** → **Surveys and overviews**; • **Social and professional topics** → *Employment issues*; Geographic characteristics;

Additional Key Words and Phrases: Crowdsourcing, crowdworkers, motivation, self-determination theory, scale, validation, invariance

Authors' addresses: L. Posch, GESIS–Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, Cologne, 50667, Germany, Graz University of Technology, Inffeldgasse 16c, Graz, 8010, Austria; email: lisa.posch@gesis.org; A. Bleier and F. Flöck, GESIS–Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, Cologne, 50667, Germany; emails: {arnim.bleier, fabian.floeck}@gesis.org; C. M. Lechner, GESIS–Leibniz Institute for the Social Sciences, B2,1, Mannheim, 68159, Germany; email: clemens.lechner@gesis.org; D. Danner, University of Applied Labour Studies, Seckenheimer Landstraße 16, Mannheim, 68163, Germany; email: daniel.danner@arbeitsagentur.de; M. Strohmaier, RWTH Aachen University, Theaterplatz 14, Aachen, 52062, Germany, GESIS–Leibniz Institute for the Social Sciences, Unter Sachsenhausen 6-8, Cologne, 50667, Germany; email: markus.strohmaier@humtec.rwth-aachen.de.

1 INTRODUCTION

During the past decade, crowd employment has emerged as a new form of short-term and flexible employment. As such, crowd employment is part of a wider trend in industrial societies toward increasingly flexible work arrangements that are characterized by short-term, market-based contracts [41, 50]. Crowd employment has been defined as a type of employment that "uses an online platform to enable organisations or individuals to access an indefinite and unknown group of other organisations or individuals to solve specific problems or to provide specific services or products in exchange for payment" [59]. While this definition is similar to the concept of crowdsourcing [47], it explicitly includes only those activities that are performed in exchange for payment.

One type of crowd employment platforms are microtask platforms such as Amazon Mechanical Turk[1] (AMT) or CrowdFlower.[2] On microtask platforms, crowdworkers are paid on a per-task basis, and a single task usually pays only a few cents upon completion. The microtasks offered to workers on these platforms are also called "human intelligence tasks" and typically require workers to solve problems that are easy to solve for humans but hard to solve for computers. This characteristic led Amazon Mechanical Turk to coin the term "artificial artificial intelligence" to describe this type of work. Typical microtasks include classification and tagging of text or images, audio and image transcription, and validating addresses of companies on the web. Also more complex tasks such as editing text documents [5], ontology alignment [81] and the evaluation of unsupervised machine learning algorithms (e.g., References [10, 31, 74]) have been successfully deployed on microtask platforms. Anyone, regardless of geographical location or education, can perform microtasks—the only necessary requirement is having access to the Internet.

The emergence of crowd employment and a general trend towards more flexible and shorter-term employment have given rise to policy discussions on social protection and working conditions of crowdworkers (e.g., References [15, 16, 27, 72]). One ongoing discussion is whether crowd employment is to be considered "work" at all, or whether it is mostly considered a spare-time activity by many workers, meaning that remuneration plays only a minor role for them [72]. Estimates of the hourly wage achievable on popular microtask platforms lie between under US$1 and around US$5 [3, 4, 44, 52, 77]. While this amount is above the minimum wage in some countries, in many high-income countries it is far below the wage of any traditional job. Despite this, the rise of crowd employment is an international phenomenon that does not exclude high-income countries. Understanding the underlying motivations of the international, "indefinite and unknown group" of crowdworkers is therefore crucial for understanding this new form of employment.

Although there has been some research on the motivation of crowdworkers, there is currently no theoretically founded scale for comprehensively measuring motivations of crowdworkers in different countries. So far, the question of what motivates people across the world to participate in microtask crowdwork remains largely open.

This article lays the groundwork for understanding the motivations of the international crowd workforce by introducing the Multidimensional Crowdworker Motivation Scale (MCMS) and

---

[1]http://www.mturk.com/.
[2]http://www.crowdflower.com/.

presenting a case study conducted on a large sample of crowdworkers from ten different countries. The MCMS is theoretically grounded in self-determination theory (SDT) and tailored specifically to the context of paid crowdsourced microlabor. Most items in the MCMS are based on items from existing SDT-based motivation scales developed for the traditional work context, which we adapted to the idiosyncrasies of work on microtask platforms.

The main contributions of this article are (1) an evaluation of two existing SDT-based work motivation scales developed for the traditional work context with respect to their suitability for microtask crowdwork; (2) the development of the Multidimensional Crowdworker Motivation Scale (MCMS) that draws from these existing scales but refines them and adapts them to the context of microtasks; (3) a validation of the MCMS in ten countries and three income groups; (4) an evaluation of the comparability of motivations measured with the MCMS across countries and across income groups; and (5) a first analysis of differences in crowdworker motivations across these countries and income groups. To the best of our knowledge, the MCMS is the first motivation scale developed specifically for the context of crowdsourced microlabor that offers a comprehensive representation of the motivational dimensions according to SDT. Furthermore, it is the first motivation scale for the crowdworking domain that is validated across multiple countries and income groups.

The article is structured in the following way: Section 2 gives a short overview of the different types of motivation as conceptualized by self-determination theory and reviews existing SDT-based work motivation scales. Furthermore, it gives an overview of related work on the motivations of crowdworkers on microtask platforms. In Section 3, we evaluate to what extent SDT-based motivation scales developed for the traditional work context can be successfully applied for measuring crowdworker motivations, and we show the need for a work motivation scale adapted to the idiosyncrasies of the microtask context. Section 4 describes the process of developing the MCMS, and Section 5 presents a validation of the MCMS in ten countries and three income groups. In Section 6, we demonstrate the cross-country and cross-income group comparability of motivations measured with the MCMS. Section 7 presents a first cross-country and cross-income group comparison of crowdworker motivations. Finally, Section 8 concludes this work and discusses the scale's limitations as well as directions for future research.

2 RELATED WORK

Self-Determination Theory and Work Motivation. Self-determination theory (SDT) is a theory of human motivation that was developed by Deci and Ryan [20–22]. The theory specifies three general kinds of motivation that are hypothesized to lie along a continuum of self-determination: *amotivation*, *extrinsic motivation* and *intrinsic motivation*. At the one extreme of the continuum lies *amotivation*, which completely lacks self-determination; at the other extreme lies *intrinsic motivation*, which is completely self-determined [33]. Between these extremes lies *extrinsic motivation*, which is further split up into subtypes with varying degrees of internalisation: *external regulation*, *introjected regulation*, *identified regulation* and *integrated regulation*.

Figure 1 (based on Reference [33]) shows the types of motivation as specified by SDT. *Amotivation* is the absence of motivation, a state of acting passively or not intending to act all. *External regulation* is the least self-determined form of extrinsic motivation. Individuals motivated by external regulation act to obtain rewards or avoid punishments. *Introjected regulation* refers to a form of partially internalized extrinsic motivation that aims at the avoidance of guilt or at attaining feelings of worth [24]. *Identified regulation* is a form of extrinsic motivation with a high degree of perceived autonomy, where the action is in alignment with the individual's personal goals. *Integrated regulation* is the most self-determined form of extrinsic motivation and stems from evaluated identifications that are in alignment with self-endorsed values, goals and needs
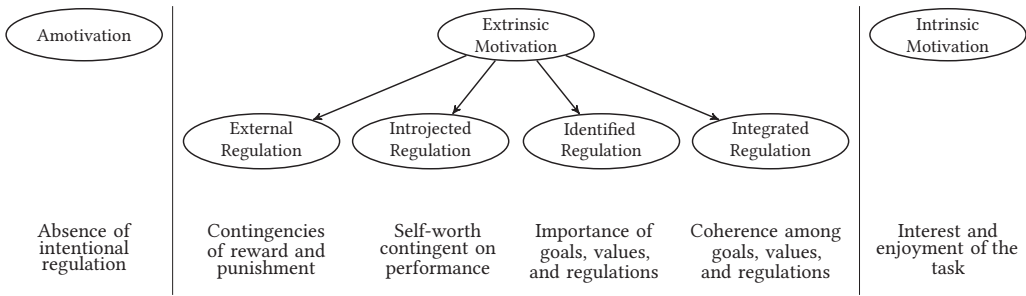
Fig. 1. **Types of motivation.** This figure shows the different types of motivation along the self-determination continuum hypothesized by SDT. The figure is based on Gagné and Deci [33].

[24]. The most self-determined form of motivation is *intrinsic motivation.* This form of motivation is non-instrumental and people act freely, driven by interest and enjoyment inherent in the action [80].

SDT hypothesizes that individuals may internalize an initially external regulation, which then becomes more self-determined. Regulations can be internalized in different ways, depending on the extent to which the individual has integrated it with his or her sense of self [23]. For example, an activity could initially be externally regulated, because it is not perceived as enjoyable by the individual. However, when this activity becomes valued by the person, for example, because it is perceived to be important for his or her personal goals, the regulation for this behavior is internalized (in this case, becoming identified regulation).

While SDT postulates that the different types of internalization fall along a continuum structure, empirical evidence for this continuum hypothesis is inconsistent (e.g., References [11, 37, 46, 58]). For example, Chemolli and Gagné [11] showed that motivations differ more in *kind* than in *degree*, and that SDT-based motivation scales are best represented by multidimensional models. Howard et al. [46] found evidence of a global factor measuring the quantity of self-determination, but they also found that each of the motivation types provided unique information beyond the quantity of self-determination.

Several work motivation scales for the traditional employment context have been developed based on SDT. The first SDT-based work motivation scale was a French scale developed by Blais et al. [6]. Tremblay et al. [88] translated this scale into English and conducted an evaluation in different work environments. The resulting Work Extrinsic and Intrinsic Motivation Scale (WEIMS) measures six factors: amotivation, the four external regulation subtypes and intrinsic motivation. Gagné et al. [34] created the Motivation at Work Scale (MAWS), a scale that measures the four factors external regulation, introjected regulation, identified regulation and intrinsic motivation. The MAWS was validated in French and in English and was partly based on the scale developed by Blais et al. [6].

Later, Gagné et al. [35] developed the Multidimensional Work Motivation Scale (MWMS). The MWMS was validated in seven languages and nine countries and does not include any items from the MAWS. The MWMS measures six first-order factors (amotivation, material external regulation, social external regulation, introjected regulation, identified regulation, and intrinsic motivation) and one second-order factor (external regulation). Work motivation scales such as MWMS, MAWS, and WEIMS investigate motivations at the domain level of analysis, meaning that they measure the general motivation to perform a job as opposed to specific tasks within a job.

Crowdworker Motivation on Microtask Platforms. Compared to work motivation in the traditional employment context, research on motivation in the microtask context is still scarce

and scattered. Most research investigating the motivations of workers on microtask platforms has focused on the platform Amazon Mechanical Turk (AMT). Consequently, most studies have focused on American and Indian crowdworkers, which constitute the vast majority of workers on AMT[3] [48, 49, 77]. This country distribution is likely due to the fact that workers can receive money from AMT in the USA and in India while workers from other countries are paid in Amazon.com gift cards [89].

Studies that investigated crowdworker motivation suggest that there are different motivations for participating on microtask platforms. For example, one early study on the reasons crowdworkers have for participating on AMT was conducted by Ipeirotis [49]. In this study, the author asked the multiple-choice question "Why do you complete tasks in Mechanical Turk?", offering six response options. He found that more Indians than Americans treat AMT as a primary source of income, and that few Indian workers report the reason "To kill time." Hossain [45] created a classification of motivation in online platform participation, listing extrinsic and intrinsic motivators and incentives.

Kaufmann et al. [51] developed an early model for measuring crowdworker motivations on AMT, differentiating between enjoyment-based motivation, community-based motivation, immediate payoffs, delayed payoffs, and social motivation. They used a sample composed of Indian and US workers on AMT and found that the construct with the highest score was "immediate payoffs," i.e., payment. Their study further found that the pastime score correlated positively with household income and negatively with the weekly time spent on AMT, and that workers who spend a lot of time on AMT may be motivated differently than workers who spend little time on the platform.

Antin and Shaw [1] used a list experiment to investigate social desirability effects in motivation self-reports of crowdworkers from the USA and India on AMT. Using the four items "to kill time," "to make extra money," "for fun," and "because it gives me a sense of purpose," they found that U.S. workers tended to over-report all four reasons while Indian workers tended to over-report "sense of purpose" and under-report "killing time" and "fun."

For measuring extrinsic motivations of crowdworkers, Naderi et al. [65] evaluated a four-factor model using a subset of WEIMS items on a sample of U.S. workers on AMT. In this model, identified and integrated regulation are merged into one factor, and the intrinsic motivation factor is omitted. After a first version of the present study was published [75], Naderi adapted and extended their scale to include intrinsic motivation [64]. The adapted scale, named the Crowdwork Motivation Scale, measures five constructs, three of which are measured by WEIMS items (amotivation, external regulation and identified regulation). The scale was evaluated on three samples of workers on AMT (N = 170, 90, and 86).

In addition to these few quantitative studies, several qualitative studies on the motivations of crowdworkers have been conducted. For example, Gupta et al. [38, 39] investigated, among other aspects, the motivations of Indian crowdworkers on AMT and Martin et al. [61] studied the content of a forum for AMT users. Other research related to the motivations of crowdworkers includes measuring the impact of motivation on performance [76] and manipulating motivations via task framing [9] or achievement feedback [57].

In sum, these studies demonstrate that there are meaningful differences in crowdworkers' motivations to participate on microtask platforms. However, systematic and theory-driven inquiries into the motivations of crowdworkers remain in short supply. SDT-based work motivation scales may offer a suitable foundation for such inquiries, but the applicability of such scales to the microtask domain has yet to be established.

---

[3]American and Indian crowdworkers currently constitute over 80% of the worker population on ATM (also see http://demographics.mturk-tracker.com/#/countries/all).

3   SUITABILITY OF EXISTING WORK MOTIVATION SCALES

Crowd employment on microtask platforms is similar to traditional employment in the sense that in both contexts, workers provide a service in exchange for payment. In both contexts, tasks need to be completed, and these tasks are usually specified by the employer/requester and executed by the employee/crowdworker. However, several fundamental aspects of work on microtask platforms differ from the traditional employment context. For example, on microtask platforms, the relationship between the requester and the worker is often completely anonymous and extremely short-lived, often lasting only a few minutes. Furthermore, there is only a minimal amount of communication between the requester and the workers, often not exceeding the static one-way communication via written task instructions. Regarding the social environment, there is often no communication or collaboration between co-crowdworkers, as microtasks are intended to be completed as individual work.

To determine to what extent existing SDT-based work motivation scales that were developed for the traditional work context are suitable for application in the microtask context, we conducted an evaluation of two work motivation scales, the WEIMS [88] and the MWMS [35], with crowdworkers on CrowdFlower.[4] The microtask market is dominated by two platforms, CrowdFlower and AMT, which are estimated to share 80% of all revenue generated in the microtask market, with revenues being approximately equal [55]. Our reason for choosing CrowdFlower over AMT is that we aim to provide a motivation scale suitable for an international comparison of crowdworker motivations, instead of exclusively focusing on crowdworkers based in the USA and in India. We consider CrowdFlower to be better suited for this task as it pays workers via independent partner channels[5] and therefore attracts a more international crowd-workforce.

For the scale stems and items of the WEIMS and the MWMS to be conceptually applicable to the crowdworking domain, we had to make minimal adaptations to the scales before evaluating them in the microtask context. For WEIMS, we changed the stem "Why do you do your work?" to "Why do you do CrowdFlower tasks?"[6] and replaced the word "it" (referring to "your work") in the items with "CrowdFlower tasks." The stem of MWMS "Why do you or would you put efforts into your current job?" was changed to "Why do you or would you put efforts into CrowdFlower tasks?" and words in the items referring to "your current job" were replaced with "CrowdFlower tasks." Additionally, one item in the MWMS was conceptually not applicable to the domain and had to be adapted. There is no equivalent to "losing one's job" on microtask platforms. The closest concept on CrowdFlower is failing many quality control questions, which results in a lower worker account accuracy and consequently in less tasks being offered to the worker. Therefore, the item "Because I risk losing my job if I don't put enough effort in it." was changed to "Because I risk not being offered enough tasks if I don't put enough effort into them."

Respondents answered both scales along a 7-point Likert-type scale. We adopted the verbal descriptions of the scale's endpoints from the original scales: For the adapted WEIMS (A-WEIMS), the scale ranged from "does not correspond at all" (1) to "corresponds exactly" (7) and for the adapted MWMS (A-MWMS), the scale ranged from "not at all" (1) to "completely" (7).

For both the A-WEIMS and the A-MWMS, we collected responses from 500 crowdworkers residing in the USA. The surveys were posted as tasks on CrowdFlower and anonymity was ensured

---

[4]CrowdFlower changed its name to Figure Eight in February 2018 [8].

[5]http://www.crowdflower.com/labor-channels/.

[6]We chose to use the term "CrowdFlower tasks" in the stem questions and in the items instead of a more general term, because workers who are logged into CrowdFlower via the partner channels see that they are doing "CrowdFlower tasks." We can therefore assume that workers know what CrowdFlower tasks are. In contrast, general terms like "microtasks" are widely used in scientific publications and sometimes in the media, but we cannot be sure that workers on CrowdFlower understand this term as it does not appear frequently on partner channel websites or on the platform itself.

in the description of the tasks. After removing spammers (also see Section 5), the sample size was 424 for the A-WEIMS and 414 for the A-MWMS. This entails a subject-to-item ratio higher than 20:1, which is a suitable ratio for factor analysis [32, 70]. The questionnaires adhered to the default design of the CrowdFlower platform and the design was very similar to that used in the later validation of the MCMS scale (see figures in Appendix C).

Confirmatory Factor Analysis. Confirmatory factor analysis (CFA) is a multivariate data analysis technique used to test how well latent constructs, specified according to theory, represent reality according to the data gathered [40]. We used CFA to evaluate the psychometric quality (i.e., the quality of the measurement of the latent constructs) of the A-WEIMS and the A-MWMS. In particular, we evaluated the factor structure of the items with CFAs, using the R packages lavaan [78] and semtools [83].

Because of non-normality of the item distributions, we used a robust maximum likelihood estimator (as suggested in, e.g., References [19, 28]). By specifying a robust maximum likelihood estimator, the model parameters were estimated with robust standard errors, and a Satorra-Bentler (S-B) scaled test statistic is reported [78, 82].

In the context of CFA, the validity of a measurement model is evaluated via a range of goodness-of-fit (GOF) measures. These GOF measures assess the extent to which the theory (as specified in the model) represents reality (i.e., the data). Once acceptable levels of GOF measures are established, other aspects of construct validity can be evaluated [40]. For a further discussion of construct validity, see Section 5. We evaluated the model fit based on the absolute[7] fit measures *root mean squared error of approximation* (RMSEA) and *standardised root mean square residual* (SRMR) as well as the incremental[8] fit measures *comparative fit index* (CFI) and *Tucker-Lewis index* (TLI). In line with current conventions for judging model fit (e.g., References [53, 60]), we chiefly relied on the CFI, RMSEA and and SRMR to assess model fit[9] and judged model fit to be acceptable according to the following criteria: A well-fitting model should have an RMSEA of less than 0.06, a SRMR of less than 0.08, and a CFI and TLI higher than 0.95 (but at least 0.9 to be acceptable). Furthermore, we report the (S-B scaled) Chi-Square test statistic but do not rely on it for determining model fit as it is very sensitive to sample size (e.g., Reference [2]).

Measurement Models. We tested the following models for the adapted WEIMS: (1) The original WEIMS model with six factors (A-WEIMS-M1); (2) an alternative five-factor WEIMS model with identified regulation and integrated regulation loading onto a single factor (A-WEIMS-M2); and (3) the four-factor, 12-item subset of WEIMS items used by Naderi et al. [65] to measure the extrinsic motivations of workers on Amazon Mechanical Turk (A-WEIMS-M3).[10] Our rationale for evaluating the alternative model A-WEIMS-M2 is that the integrated regulation factor has been shown to be poorly separable from identified regulation and intrinsic motivation (e.g., References [88, 90]), which is also one of the reasons for why the MWMS does not include an integrated regulation factor [35].

For the A-MWMS, we tested the model originally hypothesized by Gagné et al. [35] (A-MWMS-M1) and the model that had the best fit in Reference [35] (A-MWMS-M2). Furthermore, we tested a six-factor model in which material external and social regulation are separate factors, omitting the second-order external regulation factor. We tested this model with a hypothesized correlation of

---

[7]Absolute GOF measures measure how well the model represents the data, independently of other, alternative, models [40].
[8]Incremental GOF measures compare the specified model with a baseline model where all variables are uncorrelated [40].
[9]A discussion on the guidelines for determining model fit can be found in Hooper et al. [42].
[10]We did not include the extended and adapted version of the 12-item subset of WEIMS items [64] in the evaluation as it was published a year after the development and validation of the MCMS was completed and made available as a first version [75].

Table 1. Evaluation of Existing Work Motivation Scales

Scale/Model	N	S-Bχ^2	df	CFI	TLI	RMSEA	RMSEA 90% CI		SRMR
A-WEIMS-M2	424	500.58	125	0.908	0.888	0.084	0.077	0.091	0.067
A-WEIMS-M3	424	110.57	48	0.969	0.957	0.055	0.043	0.068	0.043
A-MWMS-M1	414	828.27	143	0.818	0.782	0.108	0.101	0.114	0.170
A-MWMS-M3	414	667.90	139	0.859	0.827	0.096	0.089	0.103	0.155
A-MWMS-M4	414	532.55	137	0.895	0.869	0.084	0.077	0.091	0.087

This table shows the goodness-of-fit measures for the different models based on existing SDT-based work motivation scales that were minimally adapted to the crowdworking domain.

zero between intrinsic motivation and both external regulation factors (A-MWMS-M3) as well as without this correlation restriction (A-MWMS-M4).

Results. Table 1 shows the goodness-of-fit statistics of the different models. None of the evaluated models, with the exception of the four-factor model A-WEIMS-M3, which does not measure intrinsic motivation or social external regulation, had an acceptable model fit on our data.

The A-WEIMS model with six factors (A-WEIMS-M1) could not be estimated due to the factors identified regulation and integrated regulation not being distinguishable from another, which resulted in the covariance matrix of the factors not being positive definite. This is consistent with the findings of Naderi et al. [65]. Also for the alternative five-factor model A-WEIMS-M2, the goodness-of-fit measures were outside the acceptable range.

The four-factor model A-WEIMS-M3 was the only evaluated model with a good fit.[11] However, besides the drawbacks that the model does not measure intrinsic motivation or social external regulation, and two of the four factors are measured by only two items each, it has additional limitations: It includes items for measuring the amotivation construct that were criticized by Gagné et al. [35] for resembling low satisfaction of the need for competence rather than measuring amotivation (e.g., "I ask myself this question, I don't seem to be able to manage the important tasks related to this work."). Besides this criticism regarding content validity (i.e., regarding the extent to which the items correspond to the conceptual definition of the construct [40]), the amotivation construct also had an average variance extracted (AVE) below 0.5 (also see Section 5), suggesting low convergent validity [40].

An examination of the estimated correlations of A-WEIMS-M3 reveals that the second-strongest positive correlation (0.38) is between the amotivation construct and the identified/integrated regulation construct.[12] This questions nomological validity (i.e., whether the relationships between the constructs correspond to theory and prior research [40]) as it is inconsistent with both SDT and the correlation patterns reported by other SDT-based motivation scales, which report a low to moderate negative correlation between these constructs (e.g., References [26, 35, 67, 88]).

Of the evaluated A-MWMS models, the six-factor model with separate factors for social external and material external regulation and no correlation restrictions had the best fit. However, the fit was still not acceptable, with all goodness-of-fit measures falling outside acceptable ranges. The fit measures for A-MWMS-M2 are not included in Table 1, because the covariance matrix of the factors was not positive definite. In sum, these results suggest that the existing scales do not allow to measure the crowdworker motivation validly. We thus decided to conduct an in-depth item analysis.

---

[11]Note, however, that this evaluated subset of A-WEIMS is not identical to the one used by Naderi et al. [65] as a result of the adaptations made in item wording. Naderi et al. [65] reported a lower CFI (0.931) and a higher RMSEA (0.088).
[12]Naderi et al. [65] reported an even higher correlation of 0.48.

To identify which items in A-WEIMS and A-MWMS were most problematic in the microtask context, we conducted an exploratory factor analysis (EFA) on both samples. In contrast to CFA, factors in EFA are not specified according to theory, but derived from the data [40]. Therefore, this technique is useful for investigating the underlying structure of a set of items, using the gathered data as a starting point. For all exploratory factor analyses, we used oblique rotation (promax), because—in line with SDT—we expected the factors (i.e., motivational dimensions) to correlate.

For the A-WEIMS, we conducted an EFA with four factors: We removed the amotivation items due to the problems described above and, because of the CFA results, we expected the integrated regulation items and the identified regulation items to load onto a single factor.[13] The most problematic remaining items were "Because this type of work provides me with security," which had the highest loading on the the same factor as the items from the identified/integrated regulation factor and "Because I want to be a 'winner' in life," which did not have a high loading on any factor. This item was also criticized by Gagné et al. [35] for being culturally sensitive. Removing these items yielded a solution with four factors (material external regulation, introjected regulation, identified/integrated regulation and intrinsic motivation), whereby two of the factors only had two items remaining.

For the A-MWMS, we conducted an EFA with six factors.[14] Again, one of the most problematic items was related to security ("Because it gives me greater job security if I put enough efforts into doing CrowdFlower tasks"), which did not load onto one factor with the other items from the material external regulation factor, but instead had a high loading on the same factor as the items from the identified regulation construct. Also two items from the introjected regulation construct ("Because I have to prove to myself that I can." and "Because it makes me feel proud of myself.") had moderately high loadings (0.57 and 0.58) on this factor. After removing these three items, the introjected regulation construct as well as several other items were still problematic. Iteratively removing all problematic items yielded a solution with amotivation measured by two items, external material regulation measured by only one item, and either no introjected regulation factor or one that was very poorly distinguishable from the social external regulation factor.

Thus, our results show that the evaluated work motivation scales developed for the traditional work context do not work well within the crowdworking context when only minimal adaptations in item wording are made. While both scales contain items that are potentially useful for measuring the motivations of crowdworkers, neither of the scales is an accurate measure of the full spectrum of motivational dimensions in the microtask context. For the development of a reliable motivation scale that fills these gaps and measures the motivations of crowdworkers on all dimensions proposed by SDT, further adaptations are needed.

4 DEVELOPMENT OF THE MCMS

The results of our factor analyses conducted on the slightly modified WEIMS and MWMS underscore the necessity for developing a new scale for measuring the motivations of crowdworkers that is adapted to the idiosyncrasies of the crowdwork environment. To meet this necessity, we developed the Multidimensional Crowdworker Motivation Scale (MCMS). The MCMS was developed to provide a psychometrically sound scale that covers the motivational dimensions proposed by self-determination theory and that can be answered by crowdworkers in a limited amount of time.

---

[13]An EFA with five factors showed that one item from the identified regulation construct and one item from the integrated regulation construct loaded onto a separate factor, while all other items from these constructs loaded onto a single factor. Therefore, we proceeded with the four-factor version.

[14]An EFA with five factors showed that the items from the material and social external regulation constructs did not load onto the same factor.

We proceeded in three steps. First, we compiled a pool of items conceptually suitable for the characteristics of the crowdworking domain. An item pool is a set of candidate items that reflect the latent constructs the scale intends to measure. During the development of a scale, this set of candidate items is then reduced and refined by deleting items that exhibit undesirable properties such as high loadings on multiple factors or no high loadings on any factor, with the goal of arriving at a final scale with optimized reliability and scale length (e.g., References [25, 93]). Second, we thus selected items from the pool based on an exploratory factor analysis on a sample of workers residing in the USA. Third, we further reduced and refined the item pool based on exploratory factor analyses on samples from Spanish[15] and Indian crowdworkers.

Item Pool Generation. The results of our evaluation described in the previous section indicate that both the A-WEIMS and the A-MWMS contain items that are potentially useful for application in the microtask context. Therefore, for compiling the item pool, we first included all items from the A-WEIMS[16] and the A-MWMS that were not among the most problematic items according to the results of the factor analyses. We retained the adaptations to item wording. To extend the pool, we added semantically suitable items from the SDT-based motivation scales developed in References [67] and [26] as well as nine new items developed by the authors.[17] The total number of items in the pool was 44.

To ensure content validity (i.e., the extent to which the items correspond to the theoretical constructs [40]), we based the majority of candidate items on existing scales, and we closely followed the definitions of the constructs during the selection of candidate items from existing scales as well as during the creation of the new items. We used the construct definitions provided in the Handbook of Self-Determination Theory Research [23] and a publication by Gagné and Deci discussing the constructs in the context of work motivation [33].

We used the stem phrasing "Why do you or would you put efforts into doing CrowdFlower tasks?", adapted from MWMS, to capture both actual and latent motivations. The items were rated along a 7-point Likert-type scale ranging from "not at all" (1) to "completely" (7).[18]

First Round of Data Collection. For a first selection of items from the pool, we collected answers from 1,000 crowdworkers residing in the USA and conducted an exploratory factor analysis on their responses. Consistent with the findings from Section 3, we found that the items intended to measure material external regulation and social external regulation did not load on the same factor. Therefore, we aimed at a six-factor model with the two separate external regulation factors (social and material). Items that had insufficient loadings on the appropriate factor (<0.5), items that loaded on a factor other than the hypothesized one, and items with high cross-loadings (>0.35) were iteratively removed from the initial pool, creating a reduced item pool with 36 items.

Second Round of Data Collection. We conducted a second round of data collection with this reduced item pool, collecting responses from 1,000 Spanish and 1,200 Indian crowdworkers.[19] The additional 200 responses from the Indian crowdworkers were requested because of the high

---

[15]In this article, we use country demonyms synonymously with the location of workers for better readability.

[16]Notably, this also included all items from A-WEIMS-M3 except for the amotivation items due to the problems described in Section 3.

[17]The development of the new items focused mainly on the material external regulation construct, as material external regulation is a construct present only in scales intended to measure motivation in a context where material rewards (such as money) are relevant. It is therefore not present in most SDT-based motivation scales.

[18]The verbal descriptions for each scale point were adopted from the MWMS [35] and shown to the participants in the task instructions: 1 = "not at all," 2 = "very little," 3 = "a little," 4 = "moderately," 5 = "strongly," 6 = "very strongly," 7 = "completely."

[19]India was selected, because it has a significant population of crowdworkers on different platforms. Spain was selected to include a European country with a significant population of crowdworkers.

amount of spam received in this group (also see Section 5), with the aim of achieving an item-to-response ratio of close to 1:20. Again, we iteratively removed items with low loadings (with the higher threshold of <0.7 if more than three items were left for this construct) or high cross-loadings (with a threshold of >0.3). Furthermore, if two items were phrased very similarly and the factor had more than three items remaining, we removed the item with the lower loading.

The final MCMS contains 18 items, with three items measuring each factor. Of the 18 final items, five items (Am2, Introj2, Introj3, Intrin1, Intrin3) were adapted from Reference [35], four items (ExMat2, Ident1, Ident2, Ident3) from Reference [88], two items (Am3, Introj1) from Reference [26], two items (ExSoc2, Intrin2) from Reference [67], and five items (Am1, ExMat1, ExMat3, ExSoc1, ExSoc3) are new (but semantically based on items from existing scales). Like other SDT-based work motivation scales such as the MWMS [35], the MCMS aims to measure motivations for putting effort into the job (in this case microtasks) in general, as opposed to measuring the motivations for specific tasks within a job. Table 14 in Appendix A shows the scale.

5 VALIDATION OF THE MCMS

Data Collection. With the final 18-item version of the MCMS, we collected data from 10 countries, with 900 participants from each country, for validation. We selected countries from three World Bank income groups[20]: high-income, upper-middle-income, and lower-middle-income. From each of the three income groups, we selected three countries with high activity on CrowdFlower. The countries were selected according to the following criteria: First, the country had to be active on CrowdFlower (either high in the Alexa[21] ranking or one of the top contributing countries in at least one of the partner channels). Second, we aimed for a high cultural diversity overall as well as within the income groups. For the high-income group, the selected countries were USA, Germany and Spain. The upper middle-income group contains Brazil, Russia and Mexico, and the lower middle-income group is comprised of India, Indonesia and the Philippines. Note that in the rest of this article, we use the group label "Middle Income" (MID) for the upper middle-income group and "Low Income" (LOW) for the lower-middle-income group for better readability.

In addition, we collected responses from Venezuela, because it was the most active country on CrowdFlower at the time of data collection, with CrowdFlower receiving 18.5% of traffic from this country.[22] However, we did not include Venezuela in the data grouped by income, because we believe it represents a special case: At the time of data collection, the US$ earned on CrowdFlower could be sold on the black market at a rate several orders of magnitude higher than the official exchange rate [66].

To capture a diverse sample of crowdworkers in each country, the starting times of the survey were divided into three groups: (1) 300 responses were requested during typical working hours (8:00 am to 5:00 pm in the appropriate time zone), (2) 300 responses were requested in the evening (6:00 pm to 11:00 pm in the appropriate time zone) and finally, (3) 300 responses were requested during weekends. We made the survey available to workers of all CrowdFlower levels. The data was collected in October and November 2016. A full description of the demographic characteristics of our sample can be found in Posch et al. [73] and on our website [29].

Task Interface and Payment. The items in the MCMS were randomly permuted and presented to crowdworkers as a task on CrowdFlower. Besides the MCMS items, the task also included a section with demographic questions and questions about money use, as well as a section in which workers were instructed to think of five reasons for why they do tasks on CrowdFlower and asked

---

[20]http://databank.worldbank.org/data/download/site-content/CLASS.xls.
[21]http://www.alexa.com/.
[22]Data obtained from http://www.alexa.com/.

to write down these reasons. Anonymity was guaranteed in the task instructions. The interface[23] of the task is shown in Appendix C. After completing a task, crowdworkers on CrowdFlower are asked by the platform to judge the task according to different criteria, one of them being the clarity of the task instructions and interface. The question asked is *"How clear were the task instructions and interface?"* and workers are asked to answer on a five-point scale ranging from "very unclear" to "very clear." The average of the workers' responses was higher than 4.0 in all countries except for Brazil (3.6) and Indonesia (3.8). This indicates that the task instructions and interface was perceived as clear in most countries, and as "somewhat clear" in Brazil and Indonesia.

We aimed for a payment similar to most other tasks on CrowdFlower to minimize population bias in the responses. Based on studies reporting average earnings on microtask platforms (e.g., References [3, 44, 52]) and on the first author's experience as a crowdworker on CrowdFlower, we payed US$0.1 for the task excluding platform fees. The survey presented by the platform to workers after completing a task includes a question about task payment (*"How would you rate the pay for this task relative to other tasks you've completed?"*) that crowdworkers answer on a five-point scale ranging from "much worse" to "much better." For the different countries, the averages of the workers' responses ranged from 3.5 (in Germany) to 4.1 (in Mexico and India), indicating that our task payment was equal to or "somewhat better" than other tasks according to the workers' perception.

Spam Detection. We expected a significant amount of spam in the responses, such as people not reading the questions and clicking randomly or workers accepting the task despite having insufficient English skills. To counteract a high amount of noise in the dataset, we included three test items in the motivation scale section of the CrowdFlower task, and an additional test question in the demographics section.[24] Employing test questions is a common method to implement attention checks in survey design (e.g., References [54, 69, 92]) and has also been employed in crowdsourcing research tasks (e.g., Reference [71]).

The three test items in the motivation scale section of the task asked participants to answer with a specific ranking on a 7-point scale, and the test question in the demographics section consisted of the question "Are you paying attention to the questions?" with the possible answers "No," "Yes," and "I don't know" selectable from a drop-down list. These questions ensured that less than 0.1% ($(\frac{1}{7})^3 * \frac{1}{3}$) of spammers passed the test questions, assuming that all four questions were answered at random. Table 2 shows the percentage of spam and the sample size after spam removal in each country and income group. The table also introduces the country and group codes used in the remainder of this article.

Hypothesized Model. Our hypothesized model measures six constructs and is depicted in Figure 2. The inclusion of a social external regulation construct in addition to material external regulation was adopted from the MWMS [35], because both social and material rewards are important in the work context [35, 85]. As suggested by our evaluation of existing scales and our results from exploratory factor analysis on the MCMS item pool, we modeled material external and social external regulation as two separate factors. Hence, our hypothesized model is a six-factor model in which all factors are first-order factors. Note that the material and social external regulation are not adjacent factors in the continuum hypothesized by SDT but occupy the same spot.

---

[23]English was chosen as the interface language for all countries for two reasons: First, CrowdFlower's default interface language is English and all workers are expected by the platform to understand instructions in English. This is underscored by the fact that "English" is not selectable in the requester interface when choosing crowdworkers of a specific language. Second, translating stems and items has a risk of introducing semantic mismatches. We hence weighted the risk of introducing translation mismatches higher than potential errors due to non-native speakers' misinterpretations.
[24]As CrowdFlower does not offer built-in quality control mechanisms for survey-type tasks, we did not use any platform-specific quality control mechanisms.

Table 2. Sample Sizes and Percentage of Spam Received

Group	Code	N_{raw}	Spam	N_{clean}
All	ALL	9,000	35 %	5,857
High Income	HIGH	2,700	28 %	1,952
Middle Income	MID	2,700	32 %	1,834
Low Income	LOW	2,700	44 %	1,508
USA	USA	900	20 %	721
Spain	ESP	900	25 %	677
Germany	DEU	900	38 %	554
Brazil	BRA	900	45 %	496
Russia	RUS	900	25 %	677
Mexico	MEX	900	27 %	661
India	IND	900	32 %	608
Indonesia	IDN	900	55 %	401
Philippines	PHL	900	45 %	499
Venezuela	VEN	900	37 %	563

This table shows the sample sizes of the different groups before and after spam removal, as well as the percentage of spam received.

Whereas SDT hypothesizes that intrinsic motivation does not correlate with external regulation [35, 79], Chemolli and Gagné [11] found that intrinsic motivation correlates with external regulation significantly for both the MWMS and the Academic Motivation Scale [90]. Based on these findings, we decided not to restrict these correlations to zero in our hypothesized model but to evaluate the model fit of both the correlation-restricted and the unrestricted model. No cross-loadings were hypothesized.

Descriptive Statistics and Internal Consistency. Table 3 summarizes the observed factor means as well as standard deviations for each of the countries and income groups in our data. Table 4 displays the Pearson correlations of the observed factor means. We used Cronbach's alpha statistic [17] to assess the internal consistency of the MCMS. Table 5 displays the values of alpha for each country and income group. Alpha provides an estimate of the lower bound of internal consistency. As a rule of thumb, values above 0.7 are considered acceptable, values between 0.6 and 0.7 questionable, values between 0.5 and 0.6 poor, and values below 0.5 unacceptable [36]. In most countries and groups, alpha exceeded 0.7 for each construct. Exceptions to this were the amotivation factor in Brazil and Venezuela as well as the material external regulation factor in Brazil and Indonesia with values between 0.5 and 0.7. Therefore, when interpreting results from these factors and groups, care should be taken. In addition to Cronbach's alpha, we calculated McDonald's coefficient omega [62] for assessing the reliability of the MCMS. The values for coefficient omega are shown in Appendix B. Compared to Cronbach's alpha, the values of coefficient omega are equal or slightly higher.

Confirmatory Factor Analysis. To validate the factor structure of our hypothesized model, we conducted a confirmatory factor analysis. Table 6 shows the results of the confirmatory factor analysis of the hypothesized model. The first item of each factor in Table 14 served as the marker variable (i.e., the loading of this item was fixed to one). The analysis was conducted on the entire dataset as well as on each group separately. As in Section 3, we followed current conventions for evaluating model fit (e.g., References [42, 53, 60]).

The results show that the hypothesized model had adequate fit overall as well as in all groups. The CFI was above 0.95 in all groups except Indonesia and Mexico (with 0.931 and 0.948, respectively). RMSEA was lower than 0.06 in all groups and SRMR was lower than or equal to 0.05 in all
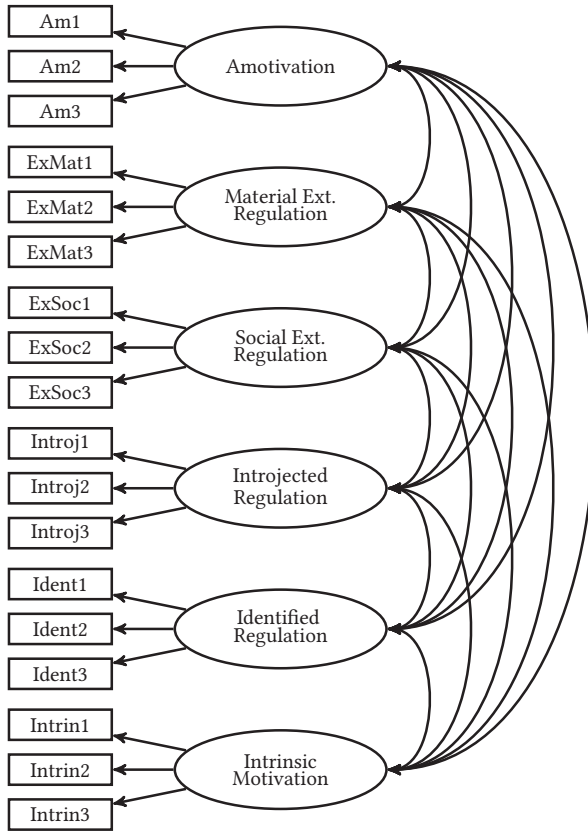
Fig. 2. **Factor structure of the MCMS.** This figure shows the factor structure of the hypothesized MCMS model. Items (i.e., measured variables) are represented by rectangles, ovals represent latent constructs, curved arrows indicate a correlational relationship between latent constructs and straight arrows indicate the dependence relationship between latent constructs and items. The individual items are shown in Appendix A.

groups. We consider the fit measures for Mexico and Indonesia to be marginally acceptable, but some care should be taken when interpreting the results from these countries. Table 7 shows the item loadings and intercepts estimated by the hypothesized model fitted to the entire sample and Table 9 shows the estimated correlations between the constructs. All estimated factor correlations are statistically significant at $p < 0.01$ (most at $p < 0.001$).

The alternative model, which restricts the correlations of the external regulation factors (material external regulation and social external regulation) with intrinsic motivation to zero, did not have an acceptable fit: While CFI was close to 0.95 for most groups (0.948 for the total sample), SRMR was high for the total sample (0.094) as well as for all other groups, ranging between 0.074 in the USA and 0.113 in Russia.

Compared to other SDT-based approaches to measuring crowdworker motivations [64, 65], the MCMS achieved a better fit while measuring additional constructs. Other SDT-based scales used to measure motivations in the microtask context [64, 65] did not report a well-fitting model for sample sizes larger than 100 due to a high RMSEA (>0.08 for the 12-item-subset of WEIMS [65] and >0.09 for the Crowdwork Motivation Scale [64]). For the Crowdwork Motivation Scale, the fit measures were better on two smaller samples (N = 90 and N = 86), but RMSEA was still high (>0.072 for

Table 3. Manifest Scale Means and Standard Deviations

Group	Amotivation	Material	Social	Introjected	Identified	Intrinsic
ALL	1.84 (1.09)	6.05 (1.08)	2.47 (1.58)	2.25 (1.46)	4.27 (1.73)	5.67 (1.26)
HIGH	2.06 (1.21)	5.86 (1.18)	1.99 (1.37)	1.91 (1.29)	3.56 (1.73)	5.27 (1.37)
MID	1.86 (1.06)	6.12 (1.02)	2.68 (1.60)	2.59 (1.58)	4.53 (1.65)	5.82 (1.18)
LOW	1.72 (1.02)	6.13 (0.99)	2.74 (1.68)	2.31 (1.44)	4.60 (1.57)	5.90 (1.11)
USA	1.91 (1.17)	5.86 (1.25)	1.67 (1.19)	1.58 (1.12)	3.38 (1.74)	5.28 (1.43)
ESP	2.10 (1.20)	5.96 (1.07)	2.46 (1.48)	2.34 (1.45)	3.89 (1.71)	5.37 (1.33)
DEU	2.23 (1.26)	5.72 (1.21)	1.82 (1.31)	1.83 (1.15)	3.40 (1.67)	5.15 (1.35)
BRA	1.89 (0.97)	6.28 (0.87)	2.59 (1.60)	2.38 (1.48)	4.62 (1.71)	5.98 (1.10)
RUS	1.95 (1.10)	5.93 (1.15)	2.67 (1.67)	2.83 (1.68)	4.48 (1.61)	5.60 (1.28)
MEX	1.75 (1.09)	6.18 (0.95)	2.76 (1.53)	2.51 (1.52)	4.51 (1.65)	5.92 (1.10)
IND	1.71 (1.02)	6.17 (1.02)	2.54 (1.67)	2.33 (1.49)	4.56 (1.62)	5.92 (1.12)
IDN	1.99 (1.16)	6.12 (0.88)	3.06 (1.70)	2.65 (1.43)	4.62 (1.40)	5.86 (1.11)
PHL	1.53 (0.84)	6.09 (1.05)	2.71 (1.63)	2.00 (1.30)	4.63 (1.63)	5.92 (1.09)
VEN	1.31 (0.62)	6.27 (0.98)	2.76 (1.56)	2.14 (1.43)	4.97 (1.66)	5.97 (1.09)

This table shows the manifest scale means and standard deviations for all groups and factors.

Table 4. Manifest Correlations between Factors

	Amotivation	Material	Social	Introjected	Identified
Material	-0.21^{***}				
Social	0.02	0.12^{***}			
Introjected	0.13^{***}	0.08^{***}	0.52^{***}		
Identified	-0.18^{***}	0.35^{***}	0.43^{***}	0.39^{***}	
Intrinsic	-0.43^{***}	0.31^{***}	0.28^{***}	0.20^{***}	0.46^{***}

This table shows the Pearson correlations between the observed scores of the six factors of the MCMS, calculated on the total sample (N = 5857). $^{***}p < 0.001$.

both samples) and CFI was below 0.95.[25] In contrast, the MCMS achieved good fit measures on the total sample and for the individual countries (with CFIs above 0.95 for most countries and an RMSEA below 0.06 for all countries).

Compared to A-WEIMS-M3, the model that had the best fit in our evaluation of traditional work motivation scales, the MCMS showed a similar fit and suffers from none of A-WEIMS-M3's drawbacks described in Section 3. Moreover, if only the four constructs measured by A-WEIMS-M3 are taken into account, then the MCMS achieves a better fit: A four-factor model of the MCMS with the intrinsic motivation and social external regulation factors omitted had a CFI of 0.983, a TLI of 0.984, an RMSEA of 0.037 and an SRMR of 0.026 on the total sample (N = 5857) and a CFI of 0.979, a TLI of 0.971, an RMSEA of 0.039 and an SRMR of 0.035 on the USA sample (N = 721).

Construct Validity. Construct validity refers to the extent to which the items of a scale "accurately reflect the theoretical latent constructs they are designed to measure" [40]. In the context of confirmatory factor analysis, a poor model fit is considered evidence of a lack of construct validity. Besides achieving a good model fit, there are four additional components of construct validity that can be evaluated: content validity, convergent validity, discriminant validity, and nomological validity [40]. Content validity was ensured during item development (see Section 4). Evidence of

---

[25]TLI and SRMR were not reported.

Table 5. Internal Consistency of the MCMS

Group	Amotivation	Material	Social	Introjected	Identified	Intrinsic
ALL	0.78 (0.77 0.79)	0.78 (0.77 0.79)	0.84 (0.83 0.84)	0.83 (0.82 0.84)	0.87 (0.86 0.88)	0.88 (0.88 0.89)
HIGH	0.84 (0.82 0.85)	0.82 (0.80 0.83)	0.85 (0.84 0.86)	0.86 (0.85 0.87)	0.87 (0.86 0.88)	0.90 (0.89 0.91)
MID	0.70 (0.68 0.73)	0.75 (0.73 0.77)	0.82 (0.81 0.84)	0.83 (0.82 0.84)	0.86 (0.85 0.87)	0.87 (0.86 0.88)
LOW	0.78 (0.76 0.80)	0.77 (0.75 0.79)	0.85 (0.84 0.87)	0.80 (0.78 0.82)	0.85 (0.83 0.86)	0.87 (0.86 0.88)
USA	0.86 (0.84 0.87)	0.84 (0.82 0.86)	0.83 (0.81 0.85)	0.83 (0.81 0.85)	0.85 (0.84 0.87)	0.90 (0.89 0.91)
ESP	0.83 (0.81 0.86)	0.80 (0.78 0.83)	0.85 (0.83 0.87)	0.87 (0.86 0.89)	0.89 (0.88 0.91)	0.90 (0.89 0.92)
DEU	0.82 (0.80 0.85)	0.80 (0.78 0.83)	0.87 (0.86 0.89)	0.82 (0.79 0.85)	0.85 (0.82 0.87)	0.86 (0.84 0.88)
BRA	0.52 (0.45 0.59)	0.67 (0.62 0.72)	0.83 (0.80 0.85)	0.81 (0.78 0.84)	0.89 (0.87 0.90)	0.86 (0.84 0.88)
RUS	0.77 (0.74 0.80)	0.81 (0.78 0.83)	0.88 (0.87 0.90)	0.88 (0.86 0.90)	0.85 (0.83 0.87)	0.89 (0.88 0.90)
MEX	0.79 (0.76 0.82)	0.72 (0.68 0.75)	0.77 (0.74 0.80)	0.78 (0.75 0.81)	0.85 (0.83 0.87)	0.86 (0.84 0.88)
IND	0.74 (0.70 0.78)	0.82 (0.79 0.84)	0.87 (0.85 0.88)	0.79 (0.76 0.82)	0.85 (0.83 0.87)	0.87 (0.86 0.89)
IDN	0.79 (0.75 0.82)	0.65 (0.59 0.71)	0.85 (0.83 0.88)	0.79 (0.75 0.82)	0.79 (0.75 0.82)	0.84 (0.82 0.87)
PHL	0.80 (0.77 0.83)	0.80 (0.77 0.83)	0.84 (0.81 0.86)	0.82 (0.79 0.84)	0.88 (0.86 0.90)	0.88 (0.87 0.90)
VEN	0.60 (0.55 0.66)	0.77 (0.74 0.80)	0.76 (0.73 0.80)	0.77 (0.73 0.80)	0.86 (0.84 0.88)	0.83 (0.81 0.86)

This table shows Cronbach's alpha values for all groups and constructs, along with a 95% confidence interval for the values.

Table 6. Confirmatory Factor Analysis of the MCMS

Group	N	S-Bχ^2	df	CFI	TLI	RMSEA	RMSEA 90% CI	SRMR
ALL	5857	1590.49	120	0.965	0.955	0.046	0.044 0.048	0.037
HIGH	1952	573.64	120	0.970	0.961	0.044	0.041 0.047	0.037
MID	1834	557.07	120	0.964	0.955	0.045	0.041 0.048	0.038
LOW	1508	554.52	120	0.955	0.942	0.049	0.045 0.053	0.036
USA	721	281.45	120	0.965	0.956	0.043	0.037 0.049	0.040
ESP	677	272.66	120	0.975	0.968	0.043	0.037 0.050	0.040
DEU	554	284.42	120	0.955	0.943	0.050	0.043 0.056	0.044
BRA	496	256.58	120	0.957	0.946	0.048	0.040 0.055	0.044
RUS	677	311.67	120	0.966	0.957	0.049	0.043 0.055	0.039
MEX	661	316.27	120	0.948	0.933	0.050	0.043 0.056	0.048
IND	608	272.66	120	0.962	0.951	0.046	0.039 0.052	0.042
IDN	401	272.08	120	0.931	0.912	0.056	0.049 0.064	0.050
PHL	499	291.47	120	0.954	0.941	0.054	0.046 0.061	0.045
VEN	563	217.07	120	0.966	0.956	0.038	0.031 0.045	0.039

This table shows the goodness-of-fit statistics for the hypothesized MCMS model. The fit statistics are given for the total sample as well as for all groups.

Table 7. Estimated Loadings and Intercepts

	Amotivation		Material		Social		Introjected		Identified		Intrinsic	
	λ	τ	λ	τ	λ	τ	λ	τ	λ	τ	λ	τ
item 1	0.775	1.824	0.823	5.987	0.894	2.285	0.787	2.251	0.829	4.274	0.868	5.623
item 2	0.745	1.683	0.627	6.101	0.832	2.041	0.750	2.309	0.828	3.967	0.817	5.758
item 3	0.700	2.012	0.776	6.059	0.710	3.086	0.826	2.186	0.845	4.560	0.855	5.637

This table shows the standardized item loadings (λ) and the item intercepts (τ) estimated by the hypothesized model (N = 5857). The first item of each construct is the marker item, with its (unstandardized) loading fixed at 1. The order of the items follows the order in Table 14.

Table 8. Variance Extracted and Shared Variance

	Amotivation	Material	Social	Introjected	Identified	Intrinsic
AVE	0.548	0.557	0.665	0.621	0.696	0.718
MSV	0.274	0.194	0.360	0.360	0.276	0.276

This table shows the average variance extracted (AVE) and maximum shared variance (MSV) between the constructs. An AVE ≥ 0.5 indicates good convergence, and an AVE > MSV is evidence for discriminant validity.

the validity of the MCMS with respect to convergent, discriminant and nomological validity is discussed below. We follow the definitions of Hair et al. [40] for these types of validity.

Convergent Validity is established if the items measuring a construct "converge or share a high proportion of variance in common" [40]. All factor loadings should be statistically significant and, especially for large samples, should be at least ≥0.5 (ideally, ≥0.7) [40]. Table 7 shows the standardized factor loadings for each construct in the MCMS. All factor loadings, except for the item ExMat2 with a loading of 0.627, were 0.7 or higher and all loadings were statistically significant at $p < 0.001$, which provides evidence of convergent validity.

In addition to inspecting the factor loadings, the average variance extracted (AVE) can be used as a summary indicator of convergence. The AVE measures the average percentage of variation in the items explained by their common latent variable. AVE is calculated as the mean of the squared standardized factor loadings of a construct. An AVE of ≥0.5 is considered to indicate good convergence [30, 40]. Table 8 shows the AVE for each construct. For all constructs, the AVE was ≥0.5, providing further evidence of convergent validity.

Discriminant Validity is "the extent to which a construct or variable is truly distinct from other constructs or variables" [40]. To establish discriminant validity, each construct's AVE should be greater than its squared correlations with other constructs, indicating that more variance in the construct's items is explained by the construct than the construct shares with other constructs [30, 40]. The highest squared correlation of a construct with any other construct, or maximum shared variance (MSV), is shown in Table 8. For all constructs, the AVE was larger than the MSV, providing evidence of discriminant validity.

Nomological Validity is concerned with whether the relationships between the constructs of the scale correspond to theory and prior research [40]. Table 9 shows the estimated correlations between the constructs of the MCMS. As hypothesized by SDT and found in previous studies (e.g., References [35, 88]), we observe a negative correlation between intrinsic motivation and amotivation (as well as a moderate negative correlation between identified regulation and amotivation). Furthermore, as hypothesized by SDT, the strongest correlations are between adjacent constructs (intrinsic motivation with identified regulation and, introjected regulation with social external regulation), and generally, constructs tend to correlate more strongly with adjacent constructs

Table 9. Estimated Correlations between Constructs

	Amotivation	Material	Social	Introjected	Identified
Material	-0.263^{***}				
Social	0.051^{**}	0.128^{***}			
Introjected	0.151^{***}	0.108^{***}	0.600^{***}		
Identified	-0.222^{***}	0.440^{***}	0.458^{***}	0.449^{***}	
Intrinsic	-0.523^{***}	0.362^{***}	0.283^{***}	0.230^{***}	0.525^{***}

This table shows the model estimates of the correlations between constructs (N = 5857). $^{**}p < 0.01$, $^{***}p < 0.001$.

than with non-adjacent constructs. There are exceptions to this pattern; notably concerning the material external regulation construct, which has a stronger correlation with intrinsic motivation and identified regulation than with adjacent constructs. However, deviations from the pattern of ordered correlations are not uncommon in SDT-based motivation scales (e.g., References [11, 46, 58, 68]). Research suggests that if a continuum structure of motivation exists, then its nature is not necessarily in line with the assumptions of SDT or represented well by the pattern of correlations estimated by CFA (e.g., References [46, 58]). Therefore, further research is necessary to determine the extent to which the MCMS follows a continuum structure and to what extent this continuum structure is in line with the nature of the continuum hypothesized by SDT. Furthermore, while SDT hypothesizes that external regulation and intrinsic motivation are unrelated, previous research found a significant positive correlation between these constructs for both the MWMS (0.11) and the Academic Motivation Scale (0.49) [11]. In line with these previous empirical results, our results also show a positive correlation between intrinsic motivation and external regulation.

Criterion Validity. For a first analysis of the relationship between crowdworkers' motivations and their behavioral outcomes, we investigated the relationships between the MCMS's motivational constructs and (1) the time taken to complete the task, (2) the amount of text content that crowdworkers produced in response to the open-ended question, and (3) the self-reported time spent on CrowdFlower per week. We used the first two measures (*time taken* and *text produced*) as an estimate for the effort that workers put into the task.

Previous empirical research found that amotivation correlates negatively with effort [35], that autonomous motivation[26] correlates more positively with effort than controlled motivation[27] [35, 84], and that autonomous motivation predicts greater effort [84]. These previous findings are consistent with SDT: For intrinsic motivation, the positive emotions associated with enjoying an activity naturally reinforce persistent effort, and identified regulation may elicit such reinforcing emotions due to value congruence even if the activity itself is not enjoyable [84]. In line with previous research [35] and consistent with SDT, we expected effort to be correlated negatively with amotivation and positively with autonomous motivation. Furthermore, we expected the correlation of effort with autonomous motivation to be more positive than the correlation of effort with controlled motivation.

Concerning the weekly time spent on the platform (*weekly time*), in line with previous research on crowdworker motivations [51], we expected the motivational profiles to differ between workers who spend a lot of time on the platform and workers who spend little time on the platform.

---

[26]Autonomous motivation encompasses identified regulation and intrinsic motivation (e.g., Reference [35, 84]).
[27]Controlled motivation encompasses external regulation and introjected regulation (e.g., Reference [35, 84]).

Table 10. Correlations between Motivational Constructs and Estimates of Effort

	Amotivation	Material	Social	Introjected	Identified	Intrinsic
time taken	-0.14***	0.04**	0.09***	0.10***	0.18***	0.15***
text produced	-0.14***	0.06***	0.00	0.03*	0.11***	0.10***
weekly time	-0.15***	0.13***	0.14***	0.08***	0.21***	0.24***

This figure shows the Pearson correlations between the different types of motivations and two estimates or effort: the time taken to complete the task and the amount of text content produced by the workers. Furthermore, it shows the Spearman correlations between the different types of motivation and weekly time spent on the platform. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Specifically, Kaufmann et al. [51] found that skill variety,[28] human capital advancement[29] and community identification[30] correlated positively with weekly time spent, while the correlation with pastime[31] was negative. Furthermore, in the context of SDT, previous research (e.g., Reference [23]) found that intrinsic motivation and highly internalized extrinsic motivation was associated with longer persistence in an activity. Therefore, we expected amotivation to be negatively correlated with weekly time spent, and we expected the strongest positive correlations to be with identified regulation and intrinsic motivation.

To calculate the time taken to complete the task, we used the starting and finishing times reported by the platform, and for an estimate of the amount of text content that a worker had produced, we counted the characters typed after removal of stopwords. Table 10 shows the Pearson correlations between the different types of motivation and the two estimates of effort.[32] While our estimates are noisy operationalizations of job effort (e.g., due to differences in reading and typing skills), they measure observed behavior as opposed to self-reports.

Our results support the hypothesis regarding effort, and we observe the same pattern as found in previous research [35] via self-reported job effort: Both measures have a significant negative correlation with amotivation and significant positive correlations with autonomous types of motivation (identified regulation and intrinsic motivation). Furthermore, the time taken correlates positively with controlled motivation types, and the amount of text produced correlates positively with material external regulation and introjected regulation. As hypothesized, the correlations between effort and autonomous motivation are more positive than the correlations between effort and controlled motivation. All differences in the strengths of the correlations between the types of autonomous motivation and the types of controlled motivation are statistically significant.[33]

---

[28]Kaufmann et al. [51] defined skill variety as "usage of a diversity of skills that are needed for solving a specific task and fit with the skill set of the worker," e.g., a worker picking "a translation task, because he likes translating." In the SDT context, this construct contains aspects of intrinsic motivation and the fulfillment of the need for competence.

[29]Kaufmann et al. [51] defined human capital advancement as "motivation through the possibility to train skills that could be useful to generate future material advantages," e.g., a worker choosing a task, "because he or she wants to improve language skills for a new or better job," which can be interpreted as an aspect of identified regulation.

[30]Kaufmann et al. [51] defined community identification as a "subconscious adoption of norms and values from the crowdsourcing platform community, which is caused by a personal identification process." In the SDT context, this construct is most similar to a regulation that has been internalized to a great extent, i.e., identified or integrated regulation.

[31]Kaufmann et al. [51] defined pastime as "acting just to kill time," e.g., a worker who "works on various 'random' tasks, because he has nothing better to do." In the context of SDT, this could be interpreted as amotivation.

[32]We logarithmized the character count as the distribution was heavily skewed.

[33]To assess the statistical significance of the differences, we used Steiger's method [86] for statistical comparisons between correlations measured on the same sample, as implemented by Lee and Preacher [56]. Our results showed that the correlations between the types of autonomous motivation and both measures of task effort were significantly stronger than the correlations between these measures and the types of controlled motivation. The differences were significant at $p < 0.01$ or lower.

To gather data on the weekly time spent on CrowdFlower, we asked workers the question *"How much time do you spend on CrowdFlower, per week?"* and workers were given seven answer possibilities, ranging from "less than 1 hour" to "more than 40 hours." This question was included in the demographics part of the task. Table 10 shows the correlations between the motivational constructs and the weekly time spent on CrowdFlower. Due to the ordinal nature of this variable, we report Spearman correlations. In line with previous research on SDT as well as previous research on crowdworker motivations, we find the highest positive correlations of weekly time spent with identified regulation and intrinsic motivation, and we find a negative correlation with amotivation.

Applicability Across Platforms. For a first evaluation of the extent to which the MCMS validly measures motivations of crowdworkers on other microtask platforms, we administered the MCMS to a small sample of crowdworkers on AMT. For use on AMT, we substituted the term "CrowdFlower tasks" with "tasks on Amazon Mechanical Turk" in the stem and items of the MCMS. We collected 150 responses from Indian workers on AMT in June 2017. After spam removal, the sample contained 109 responses. CFA results on this sample showed good fit ($CFI = 0.961$, $TLI = 0.951$, $RMSEA = 0.049$, $SRMR = 0.068$). Furthermore, measurement invariance tests (see Section 6 for more details on measurement invariance) revealed that the Indian AMT sample showed scalar invariance with our subsample of Indian crowdworkers on CrowdFlower. These results indicate that not only is the MCMS likely to be valid on other platforms, but that the MCMS also likely allows for valid comparisons of the motivational dimensions' group means across the platforms.

In sum, our results indicate that the MCMS offers reliable and valid measures of crowdworker motivations within the framework of SDT. Researchers wishing to measure the motivation of crowdworkers, for example along with other variables such as behavioral patterns of crowdworkers, can easily include the scale as a module in their task design. Instruction for use of the MCMS are given in Appendix D.

6 CROSS-GROUP COMPARABILITY OF MCMS RESULTS

When comparing the results of a measurement instrument across different groups, it is important to ensure that the instrument possesses the same psychometric properties in all groups. This characteristic is referred to as *measurement invariance*. Tests of measurement invariance evaluate "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" [43]. In our case, measurement invariance means that crowdworkers from different countries (or country income groups) assign the same meaning to the items used in the MCMS. Measurement invariance is particularly important if mean level differences between countries ought to be compared. A lack of measurement invariance can indicate, for example, that respondents of different groups understand the items in a different way (e.g., due to culture) or that different levels of response biases are present (e.g., Reference [13]).

To evaluate measurement invariance of the MCMS, we conducted multiple-group confirmatory factor analyses (MGCFA). In MGCFA, measurement invariance is evaluated via a series of hypothesis tests, which test invariance at different levels. Three levels of measurement invariance are commonly tested: configural, metric, and scalar invariance (e.g., References [12, 14]). Configural invariance requires that the items share the same configurations of loadings in all groups. Metric invariance additionally requires that the loadings of each item on its factor is the same across groups. Scalar invariance additionally requires that the intercepts of item regressions on each factor are the same across groups. To validly compare manifest mean differences across groups, scalar invariance is required. To validly compare latent mean differences across groups, at least partial scalar invariance is required [63].

We tested configural, metric and scalar invariance of the MCMS across countries and across income groups. Configural invariance is indicated by acceptable goodness-of-fit indices in an

Table 11. Measurement Invariance

	CFI	CFI Δ	RMSEA	RMSEA Δ
Income Groups				
Configural Invariance	0.964	n/a	0.046	n/a
Metric Invariance	0.963	0.001	0.045	0.001
Full Scalar Invariance	0.952	0.011	0.049	0.005
Partial Scalar Invariance*	0.955	0.008	0.048	0.004
Countries				
Configural Invariance	0.960	n/a	0.047	n/a
Metric Invariance	0.959	0.001	0.046	0.001
Full Scalar Invariance	0.930	0.028	0.058	0.011
Partial Scalar Invariance**	0.952	0.007	0.049	0.002

This table shows the results of the MGCFA tests of invariance between groups and countries. The deltas are with respect to the previous level of measurement invariance, i.e., for scalar invariance, the sum of the deltas for metric and scalar invariance should be below 0.01 for CFI and below 0.015 for RMSEA. *One free intercept (*Am3*); **five free intercepts (*Am3*, *ExMat2*, *ExSoc3*, *Introj2*, and *Ident2*).

MGCFA model without any equality constraints [91]. For indications of metric and scalar non-invariance, we follow the guidelines of Chen [12]: For metric and scalar invariance tests on large samples ($N > 300$), a change of ≥ -0.010 in CFI supplemented by a change of ≥ 0.015 in RMSEA indicates non-invariance.

Table 11 shows the goodness-of-fit indices for the progressively restricted models. The results show a good fit for the model without equality constraints, indicating that configural invariance holds. Full metric invariance was also achieved, indicating that the strength of the relationship between the items and constructs is the same across groups. Full scalar invariance could not be achieved. However, partial scalar invariance was achieved by releasing one intercept for the income groups (*Am3*) and five intercepts in the countries (*Am3*, *ExMat2*, *ExSoc3*, *Introj2*, and *Ident2*). Partial scalar invariance still allows for factor means to be compared as long as at least two intercepts per factor are invariant. Care should be taken, however, as Steinmetz' [87] simulations showed that unequal intercepts may lead to erroneous conclusions about mean-level differences when comparing observed composite means across groups. Therefore, a cross-country or cross-income group comparison of crowdworker motivations measured with the MCMS should rely on the model-implied latent means, which take intercept non-invariance into account, instead of observed composite means.

In sum, our analyses showed that the MCMS is well-suited for measuring the motivations of crowdworkers based in different countries. Furthermore, the invariance analyses indicated that the motivations measured with the MCMS can be used for a comparison of motivations across countries and across country income groups.

7 MOTIVATIONS OF CROWDWORKERS ON CROWDFLOWER

This section reports the motivations of crowdworkers on CrowdFlower as measured by the MCMS as well as the results of our cross-country and cross-income group comparison of crowdworker motivations.

Motivations Measured with the MCMS. Table 13 shows the latent means[34] of the different motivational dimensions for the entire sample (ALL). Our results show that overall, material

---

[34]To obtain estimates of the latent means, we fixed the marker items' intercepts to zero.

Table 12. Responses to the Open-ended Question

Construct	Examples
Amotivation	*"I'm bored," "I have nothing else to do," "this is boring," "Nothing."*
Material	*"I need the money," "I get paid," "To get an extra income!," "good profit."*
Social	*"other people want me to fulfill the job," "my friends do ti too," "referral from a friend," "friends do it."*
Introjected	*"I'm no worse than others," "To feel better about myself," "I feel useful," "They make me feel productive and useful."*
Identified	*"To be my own boss!," "To gain some new experience," "It helps me improve my English," "I can work from anywhere."*
Intrinsic	*"Enjoy seeing a variety of different topics," "most of the tasks are actually entertaining," "There are taske that are very interesting," "I enjoy sharing thoughts and opinions while completing CrowdFlower tasks."*

This table shows illustrative examples of answers to the open-ended question "Give five reasons for why you do CrowdFlower tasks" from crowdworkers who had a mean score of >5 on the respective construct.

external regulation was the most important motivation for crowdworkers, with a mean of 5.99, followed by intrinsic motivation with a mean of 5.62. This points to an interesting duality of monetary and interest-driven, enjoyment-based motivational influences. The result is consistent with previous research on crowdworker motivations on AMT (e.g., Reference [51]), which found that both payment and enjoyment play an important role for crowdworkers, with monetary reasons being slightly more important than enjoyment. The construct with the third highest mean was identified regulation (mean 4.27), which signifies that putting effort into CrowdFlower tasks is moderately in alignment with crowdworkers' personal goals and objectives such as lifestyle preferences or career plans. Social external regulation, introjected regulation and amotivation were the least important motivational factors for crowdworkers overall, with amotivation having the lowest score of all motivational dimensions (means 2.29, 2.25, and 1.82, respectively).

As an illustration of how the different motivational types can be interpreted in the microtask context, we give examples of the reasons that crowdworkers have for doing CrowdFlower tasks in the workers' own words. In the first section of the task, we instructed crowdworkers to think of five reasons for why they do tasks on CrowdFlower and asked them to write down these reasons. For each construct, Table 12 shows examples of answers to this question that can be interpreted to correspond to the theoretical definition of the construct. The examples given are taken, for each construct, from four different crowdworkers who had a mean score greater than 5 on the respective construct.

Differences in Motivations across Groups. As partial scalar invariance was achieved, the analysis of group differences in motivations measured with the MCMS relies on latent means estimated by the model instead of observed means. For analyzing the differences in latent constructs between groups, one group was chosen as the reference group. For our analysis, we chose the high-income group sample as the reference group for the cross-income group comparison and the USA sample as the reference group for the cross-country comparison. Figure 3 shows the differences in latent means of the different countries, compared to the reference group (USA), and Table 13 shows the latent means for all groups, as well as the mean differences to the reference group in parentheses.[35]

---

[35]We obtained the estimated differences and their statistical significances via MGCFA with the means of the reference group fixed to zero. The threshold for all reported significances is set at $p < 0.001$, except amotivation in Spain, external social regulation in Germany and external material regulation in Russia ($p < 0.05$).
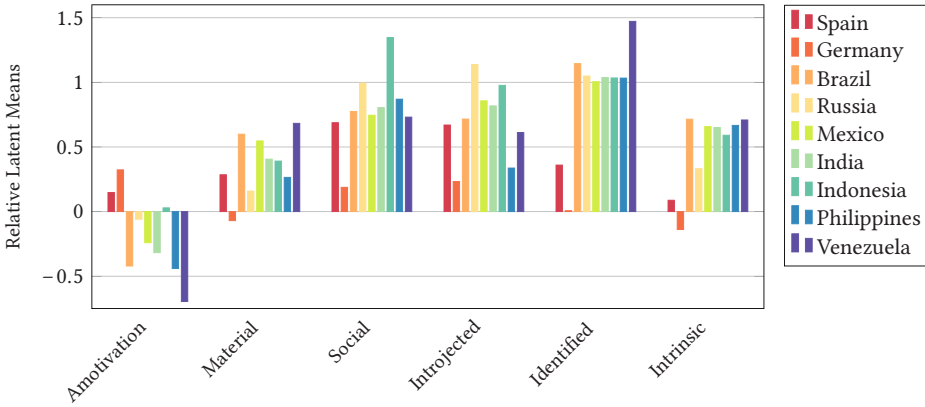
Fig. 3. **Country differences in latent means, compared to the USA sample.** This figure shows the differences in latent means for all constructs and countries, compared to the latent means of the U.S. sample.

Table 13. Latent Scale Means and Group Differences

Group	Amotivation		Material		Social		Introjected		Identified		Intrinsic	
ALL	1.82		5.99		2.29		2.25		4.27		5.62	
HIGH	2.12		5.75		1.86		1.92		3.61		5.22	
MID	1.75	(−0.36)	6.08	(0.32)	2.50	(0.64)	2.57	(0.65)	4.54	(0.93)	5.78	(0.56)
LOW	1.71	(−0.41)	6.06	(0.30)	2.58	(0.73)	2.30	(0.38)	4.60	(0.99)	5.87	(0.65)
USA	1.97		5.68		1.60		1.62		3.49		5.23	
ESP	2.12	(0.15)	5.97	(0.28)	2.29	(0.69)	2.29	(0.67)	3.85	(0.36)	5.31	(0.09)
DEU	2.30	(0.32)	5.61	(−0.07)	1.79	(0.19)	1.85	(0.23)	3.50	(0.01)	5.09	(−0.14)
BRA	1.55	(−0.42)	6.28	(0.60)	2.38	(0.78)	2.34	(0.72)	4.64	(1.15)	5.94	(0.72)
RUS	1.92	(−0.06)	5.84	(0.16)	2.60	(0.99)	2.76	(1.14)	4.54	(1.05)	5.56	(0.33)
MEX	1.73	(−0.24)	6.23	(0.55)	2.35	(0.75)	2.48	(0.86)	4.50	(1.00)	5.88	(0.66)
IND	1.66	(−0.32)	6.09	(0.41)	2.41	(0.80)	2.44	(0.82)	4.53	(1.04)	5.88	(0.65)
IDN	2.00	(0.03)	6.07	(0.39)	2.95	(1.35)	2.60	(0.98)	4.53	(1.03)	5.82	(0.59)
PHL	1.53	(−0.44)	5.95	(0.26)	2.47	(0.87)	1.96	(0.34)	4.53	(1.03)	5.89	(0.67)
VEN	1.28	(−0.70)	6.36	(0.68)	2.33	(0.73)	2.23	(0.61)	4.96	(1.47)	5.93	(0.71)

This table shows the latent means for all groups, as well as the estimated differences in latent means in parentheses.

Regarding the ranks of the constructs, the results show that the ranks of the three constructs with the highest scores were the same across all groups. External material regulation was the construct with the highest score, followed by intrinsic motivation with only a slightly lower score, in all countries and income groups. Furthermore, the third most important motivational factor was identified regulation in all groups. The other motivational dimensions differed in rank across countries and income groups. Amotivation was the construct with the lowest score in the middle- and low-income groups, while in the high-income group, it had a higher score than social external regulation and introjected regulation.

Regarding the differences in construct scores, the results show that motivations differ significantly between crowdworkers of different countries and country income groups.[36] The largest

---

[36]Note that, as described in Section 5, some care should be taken when interpreting results from Mexico and Indonesia due to a CFI lower than 0.95 (but above 0.90) and when interpreting the amotivation construct in Brazil and Venezuela, as well as the material external regulation construct in Brazil and Indonesia due to Cronbach's alpha values below 0.7.

differences in motivation, compared to workers in the USA, are with countries that are in income groups lower than the USA.

Amotivation was significantly higher in the high-income group than in the middle and low-income groups. Crowdworkers in Brazil, Mexico, India, the Philippines, and Venezuela had a significantly lower amotivation score than U.S. workers, while German and Spanish crowdworkers exhibited a significantly higher level of amotivation than workers in the USA. This indicates that in high income countries, crowdworkers tend to perceive doing CrowdFlower tasks as more "pointless" and a "waste of time" than in most lower income countries.

While material external regulation had the highest score of all constructs in all countries, scores were significantly higher in some countries than in others. Germany was the country with the lowest score on this construct, while Venezuela had the highest score. Both the middle and the low-income group reported higher scores for material external regulation than the high-income group, and crowdworkers of all countries except Germany reported a significantly higher material external regulation than U.S. workers. This means that workers in countries with lower incomes tend to be more motivated by the material rewards of microtasks than workers in high income countries.

Social external regulation and introjected regulation scores were significantly higher in the middle and low-income groups than in the high-income group, and significantly higher in all countries, compared to the USA sample. This means that satisfying external social demands, as well as the avoidance of shame or guilt feelings, are more important motivational factors for workers in low and middle income countries and in countries other than the USA.

The largest differences in construct scores, both between countries and between income groups, were found for identified regulation. Scores were significantly higher in countries of the middle and low-income groups than in countries of the high-income group. This means that in lower income countries, crowdworkers perceive putting effort into microtasks as more in line with their personal goals, objectives and values. One reason for this might be that in higher income countries, the same goals can be achieved more effectively through other means. Crowdworkers in the USA had the lowest identified regulation score of all countries and scores were significantly higher in all other countries except Germany.

Finally, all countries except Spain and Germany reported a significantly higher intrinsic motivation than workers in the USA, and middle and low-income groups reported a higher intrinsic motivation score than the high-income group. This means that that workers of countries in the middle and low-income groups are more driven by interest and enjoyment inherent in the activity.

8 CONCLUSION

In this article, we developed and validated the Multidimensional Crowdworker Motivation Scale (MCMS), a new instrument for measuring crowdworker motivations. The MCMS measures the motivations of crowdworkers on six dimensions, based on the conceptualization of motivation suggested by self-determination theory. To the best of our knowledge, the MCMS is the first instrument developed specifically for measuring motivation in the microtask context that provides a comprehensive representation of the motivational dimensions hypothesized by self-determination theory. Compared to existing instruments, the MCMS allows for a more comprehensive and theoretically well-founded measurement of crowdworker motivations with only three items per motivational dimension. Moreover, the MCMS is the first instrument for measuring crowdworker motivations that is validated in multiple countries and income groups.

The hypothesized six-dimensional model of the MCMS generally showed good fit in all countries and income groups. In addition, the results of the measurement invariance tests demonstrated that partial scalar invariance holds. This implies that the substantive meaning of the motivations measured with the MCMS are comparable across countries and income groups, allowing for valid

cross-national comparisons. By providing a reliable scale for measuring crowdworker motivations, our study constitutes an important step towards a better understanding of the international crowd-workforce.

We designed the MCMS with generalizability across platforms in mind. Our initial results on a sample of crowdworkers on AMT support the scale's applicability to other microtask platforms beyond CrowdFlower. By exchanging the platform name in the instructions and items of the MCMS, the scale can be deployed to measure the motivations of crowdworkers on other microtask platforms as well. Moreover, it likely allows for valid cross-platform comparisons of motivations, as we demonstrated with a sample of workers on AMT.

Finally, in this article, we have presented a first cross-country and income group comparison of crowdworker motivations on the microtask platform CrowdFlower. This data provides novel insights regarding wide-ranging differences of the motivations for participating on such a platform.

Given the good psychometric properties of the MCMS, future research on crowdwork can utilize the scale to address substantive questions concerning crowd employment. Here, the six motivational dimensions could serve as an outcome or as an explanatory variable. Potential fields of application[37] include predicting worker retention, investigating the relationship between worker motivation and productivity in different tasks, and conducting comparison studies of the motivations of different crowdworker populations, among others. Furthermore, the MCMS contributes to answering the question as to where in the employment space crowd employment should be located. The MCMS is also relevant for microtask platform developers who can use the scale to assess whether changes made to the platform lead to desirable or undesirable changes in motivation.

The work presented in this article has several limitations. First, the scale was presented in English to the crowdworkers in all countries, which means that we are only able to capture the motivations of crowdworkers with appropriate English skills. However, we can assume that a majority of crowdworkers on CrowdFlower possess an adequate level of English skills, as the platform interface is available exclusively in English and workers are expected to understand instructions in English. Furthermore, demand for crowdworkers is driven by English-speaking countries [55]. Regarding the presence of social desirability bias, Blais et al. [6] found that self-reported work motivations only correlated very weakly with the Marlow-Crowne Social Desirability Scale [18]. However, as Antin and Shaw [1] found evidence for the presence of social desirability bias in self-reported motivations of crowdworkers, further experiments are needed to assess the extent to which social desirability bias is present in data collected with the MCMS. A further limitation of the MCMS is that it does not measure integrated regulation. The lack of an integrated regulation factor in the MCMS is due to problems of statistically distinguishing this factor from identified regulation and intrinsic motivation. Due to the same problems, this limitation also applies to other SDT-based work motivation scales such as the MWMS [35]. Finally, the MCMS was developed specifically for the context of paid microtasks and is not intended for use in other crowdsourcing contexts. While an application in other contexts may be possible, the scale would have to be adapted first (e.g., by removing the material external regulation construct for application in the context of unpaid tasks) and validated in the respective context.

In future work, we plan to conduct a more in-depth analysis of cross-country and cross-income group differences, including their stability over time. Additionally, we plan to further investigate the relationship between motivations and economic as well as demographic factors, going beyond the country of residence as an indicator of difference. Regarding cross-platform comparability of MCMS responses, we plan to further evaluate the MCMS on other microtask platforms and,

---

[37]For instructions on use of the MCMS, see Appendix D.

provided that measurement invariance is achieved, conduct a cross-platform analysis of worker motivations.

Another direction that we plan to follow in future work is an evaluation of the extent to which motivations measured with the MCMS are related to different antecedents and outcomes. For investigating the relations between measured motivations and antecedents (e.g., the satisfaction of basic needs) or outcomes (e.g., emotional exhaustion), the scales for measuring the antecedents and outcomes will first have to be validated within the crowdworking domain. Finally, in future work, we plan to develop Bayesian models that incorporate not only the responses to the MCMS items but also responses to open-ended survey questions and demographic metadata.

To conclude, the MCMS constitutes a promising step forward in measuring the motivations of crowdworkers in a theoretically founded, reliable and internationally comparable way. By shedding light on the motivations of the "indefinite and unknown group" of crowdworkers, the MCMS enables novel insights into this emerging form of short-term employment. This work is relevant not only for researchers but also for practitioners seeking to measure the motivations of crowdworkers and to harness knowledge about differences in crowdworker motivations. Further information about our work is available on our website [29].

APPENDIX

A THE MULTIDIMENSIONAL CROWDWORKER MOTIVATION SCALE

Table 14. The Multidimensional Crowdworker Motivation Scale

		Source
Amotivation		
Am1	*I don't know why, CrowdFlower tasks often seem like a waste of time.*	
Am2	*I don't know why I'm doing CrowdFlower tasks, it's pointless work.*	[35]
Am3	*I don't know why, I often perceive CrowdFlower tasks as an annoying chore.*	[26]
External Regulation (Material)		
ExMat1	*Because CrowdFlower tasks give me financial gains.*	
ExMat2	*For the income CrowdFlower tasks provide me.*	[88]
ExMat3	*Because of the money I get from doing CrowdFlower tasks.*	
External Regulation (Social)		
ExSoc1	*Because other people want me to do CrowdFlower tasks (e.g., family, friends,...).*	
ExSoc2	*Because other people say I should (e.g., family, friends,...).*	[67]
ExSoc3	*Because other people expect it of me (e.g., family, friends,...).*	
Introjected Regulation		
Introj1	*Because otherwise I would have a bad conscience.*	[26]
Introj2	*Because otherwise I will feel ashamed of myself.*	[35]
Introj3	*Because otherwise I will feel bad about myself.*	[35]
Identified Regulation		
Ident1	*Because this is the type of work I chose to do to attain a certain lifestyle.*	[88]
Ident2	*Because I chose this type of work to attain my career goals.*	[88]
Ident3	*Because it is the type of work I have chosen to attain certain important objectives.*	[88]
Intrinsic Motivation		
Intrin1	*Because I have fun doing CrowdFlower tasks.*	[35]
Intrin2	*Because I enjoy doing CrowdFlower tasks.*	[67]
Intrin3	*Because what I do in CrowdFlower tasks is interesting.*	[35]

The stem is "Why do you or would you put efforts into doing CrowdFlower tasks?" (adapted from MWMS [35]). All items were answered on a 7-point Likert scale ranging from "not at all" (1) to "completely" (7). The column "Source" indicates from which motivation scale the item was adapted.

B MCDONALD'S COEFFICIENT OMEGA

As an alternative to Cronbach's alpha, we additionally calculated McDonald's coefficient omega [62] for assessing the reliability of the MCMS. Table 15 displays the values of omega for each country and income group. As with Cronbach's alpha, in most countries and groups, omega exceeds 0.7 for each construct. Exceptions to this are the amotivation factor in Brazil and Venezuela as well as the material external regulation factor in Brazil and Indonesia with values between 0.5 and 0.7. Compared to Cronbach's alpha, the values of coefficient omega are equal or slightly higher.[38]

Table 15. McDonald's Coefficient Omega

Group	Amotivation	Material	Social	Introjected	Identified	Intrinsic
ALL	0.78 (0.77 0.80)	0.79 (0.78 0.81)	0.84 (0.83 0.85)	0.83 (0.82 0.84)	0.87 (0.86 0.88)	0.88 (0.88 0.89)
HIGH	0.84 (0.82 0.86)	0.82 (0.80 0.84)	0.86 (0.84 0.88)	0.86 (0.84 0.88)	0.87 (0.85 0.88)	0.89 (0.88 0.91)
MID	0.71 (0.67 0.74)	0.76 (0.74 0.79)	0.83 (0.81 0.84)	0.83 (0.81 0.85)	0.86 (0.85 0.88)	0.88 (0.86 0.89)
LOW	0.78 (0.74 0.81)	0.78 (0.74 0.81)	0.86 (0.84 0.87)	0.80 (0.77 0.83)	0.85 (0.83 0.87)	0.87 (0.86 0.89)
USA	0.86 (0.83 0.88)	0.84 (0.81 0.88)	0.84 (0.80 0.88)	0.83 (0.79 0.88)	0.85 (0.83 0.88)	0.91 (0.89 0.92)
ESP	0.84 (0.81 0.87)	0.81 (0.78 0.85)	0.86 (0.83 0.88)	0.87 (0.85 0.90)	0.89 (0.88 0.91)	0.90 (0.89 0.92)
DEU	0.83 (0.8 0.86)	0.81 (0.78 0.85)	0.88 (0.85 0.91)	0.82 (0.78 0.87)	0.85 (0.82 0.87)	0.86 (0.84 0.89)
BRA	0.52 (0.44 0.61)	0.69 (0.62 0.76)	0.83 (0.80 0.86)	0.81 (0.77 0.85)	0.89 (0.87 0.91)	0.86 (0.83 0.90)
RUS	0.77 (0.73 0.81)	0.81 (0.78 0.85)	0.89 (0.87 0.91)	0.88 (0.86 0.90)	0.85 (0.83 0.88)	0.89 (0.87 0.91)
MEX	0.79 (0.74 0.84)	0.73 (0.68 0.79)	0.78 (0.74 0.81)	0.78 (0.74 0.82)	0.85 (0.83 0.87)	0.86 (0.83 0.89)
IND	0.74 (0.68 0.80)	0.82 (0.78 0.86)	0.87 (0.84 0.89)	0.79 (0.75 0.83)	0.85 (0.83 0.88)	0.88 (0.85 0.90)
IDN	0.79 (0.74 0.85)	0.65 (0.57 0.72)	0.86 (0.83 0.89)	0.79 (0.74 0.84)	0.79 (0.74 0.84)	0.85 (0.81 0.89)
PHL	0.81 (0.75 0.86)	0.80 (0.75 0.85)	0.84 (0.81 0.87)	0.82 (0.77 0.86)	0.88 (0.86 0.91)	0.89 (0.86 0.91)
VEN	0.63 (0.52 0.74)	0.77 (0.72 0.83)	0.77 (0.74 0.81)	0.77 (0.73 0.82)	0.86 (0.84 0.89)	0.84 (0.80 0.87)

This table shows McDonald's omega values for all groups and constructs, along with a 95% confidence interval for the values.

C TASK INTERFACE

The CrowdFlower task included a section in which workers were instructed to write down five reasons for why they do tasks on CrowdFlower, a section with the MCMS, and a section with questions about demographics and money use.

---

[38]One exception to this was intrinsic motivation in the high-income group, where coefficient omega is 0.01 lower than alpha.

Survey: What motivates you to do tasks on CrowdFlower?

We are interested in why you do (or would) put efforts into doing CrowdFlower tasks.

This survey consists of three parts.

Please read the instructions for each part before you complete the survey.

Instructions

Part 1: Give 5 reasons for why you do CrowdFlower tasks.

Your task in part 1: Think of 5 reasons for why you do tasks on CrowdFlower. Write one reason into each of the five text fields.

Part 2: Why do you or would you put efforts into doing CrowdFlower tasks?

Your task in part 2: Answer this survey about why you choose to put efforts into doing CrowdFlower tasks (or tasks on any other human intelligence task platform).

For each of the items, please indicate to what extent the item corresponds to the reasons for why you are putting effort into (or would put effort into) doing tasks on CrowdFlower (or on any other human intelligence task platform).

The scale for answering is:

- 1 = "not at all"
- 2 = "very little"
- 3 = "a little"
- 4 = "moderately"
- 5 = "strongly"
- 6 = "very strongly"
- 7 = "completely"

Part 3: Answer a few questions about yourself.

Your task in part 3: Please answer the questions about yourself.

This survey is anonymous and the results will be used exclusively for academic research.

Please read the questions carefully, spammers will be flagged.

Thank you very much for participating in our survey!

Fig. 4. **Task instructions.** This figure shows the task instructions that were shown to crowdworkers at the beginning of the task.

Part 1: Give 5 reasons for why you do CrowdFlower tasks.

Your task: Think of 5 reasons for why you do tasks on CrowdFlower. Write one reason into each of the five text fields.

Reason 1: (required)

Reason 2: (required)

Reason 3: (required)

Reason 4: (required)

Reason 5: (required)

Fig. 5. **Interface for open-ended answers.** This figure shows the interface that crowdworkers were given to state five reasons for why they do tasks on CrowdFlower in their own words.

Part 2: Why do you or would you put efforts into doing CrowdFlower tasks?

Using the scale below, please indicate to what extent each of the following items corresponds to the reasons why you are are putting efforts (or would put efforts) into doing CrowdFlower tasks (or tasks on any other human intelligence task platform).

Because I enjoy doing CrowdFlower tasks. (required)

	1	2	3	4	5	6	7	
not at all	○	○	○	○	○	○	○	completely

For the income CrowdFlower tasks provide me. (required)

	1	2	3	4	5	6	7	
not at all	○	○	○	○	○	○	○	completely

Because other people expect it of me (e.g. family, friends,...). (required)

	1	2	3	4	5	6	7	
not at all	○	○	○	○	○	○	○	completely

Please pick the leftmost response for this question. (required)

	1	2	3	4	5	6	7	
not at all	○	○	○	○	○	○	○	completely

Because otherwise I will feel ashamed of myself. (required)

	1	2	3	4	5	6	7	
not at all	○	○	○	○	○	○	○	completely

I don't know why, CrowdFlower tasks often seem like a waste of time. (required)

	1	2	3	4	5	6	7	
not at all	○	○	○	○	○	○	○	completely

Because CrowdFlower tasks give me financial gains. (required)

	1	2	3	4	5	6	7	
not at all	○	○	○	○	○	○	○	completely

Because I have fun doing CrowdFlower tasks. (required)

	1	2	3	4	5	6	7	
not at all	○	○	○	○	○	○	○	completely

Fig. 6. **MCMS Interface.** This figure shows the interface of the task section in which crowdworkers answered the MCMS question by indicating agreement on the 18 items. Also note the included test question to check for spamming behavior and lack of attention. Due to space limitations, this screenshot shows only part of the scale. The full scale is shown in Appendix A.

D   INSTRUCTIONS FOR USE OF THE MCMS

To measure the motivations of crowdworkers, researchers can administer the scale (shown in Appendix A) to their sample of workers. For example, the MCMS can be easily included as a module in the design of a microtask if a researcher wishes to measure motivation along with other variables, such as other characteristics or behavioral patterns, of their worker sample.

As long as the researcher does not wish to compare different groups of crowdworkers, the MCMS can be used as a summated scale. This means that the scores for the constructs can be obtained by averaging the scores of the individual items corresponding to each construct. For example, to obtain the score for the amotivation construct of a specific worker, the researcher takes the individual item responses of this worker for the items associated with the amotivation construct (i.e., Am1, Am2, and Am3) and calculates $Amotivation = (Am1 + Am2 + Am3)/3.0$.

An alternative way to obtain the construct scores is to specify the model shown in Figure 2 as a latent variable model, for example by using latent variable modeling software such as Mplus[39] or the R library lavaan.[40] This method[41] of obtaining the construct scores is preferable to averaging manifest item scores as it accounts for the measurement error necessarily present in the

---

[39]https://www.statmodel.com/.
[40]https://cran.r-project.org/web/packages/lavaan/lavaan.pdf.
[41]For an introduction to latent variable models and multivariate data analysis, we refer the reader to Kline [53], Hair et al. [40], and Bollen [7].

measurement of any abstract concept [40]. Furthermore, specifying the model allows researchers to conduct confirmatory factor analysis to validate the MCMS on their sample of crowdworkers, ensuring that the measurement is valid for their specific target population.

If a researcher wishes to compare the construct means between two groups of crowdworkers, such as male and female workers, workers of different age, workers on different platforms, or, as in the case of the present study, different countries, then the researcher has to ensure that the workers of the different groups assign the same meaning to the items used in the MCMS. This can be done by establishing measurement invariance between the groups as described in Section 6. Measurement invariance should be established before comparing the group means of any scale, especially when administering the scale to respondents of different cultures, to ensure that the differences in scores are not due to a different understanding of the items (e.g., due to culture) or due to measurement artifacts such as different levels of extreme or acquiescent response bias (e.g., Reference [13]).

REFERENCES

[1] Judd Antin and Aaron D. Shaw. 2012. Social desirability bias and self-reports of motivation: A study of Amazon mechanical turk in the U.S. and India. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI'12)*, Joseph A. Konstan, Ed H. Chi, and Kristina Höök (Eds.). ACM, 2925–2934. DOI:https://doi.org/10.1145/2207676.2208699

[2] Peter M. Bentler and Douglas G. Bonett. 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88, 3 (1980), 588.

[3] Janine Berg. 2016. Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers. *Comp. Lab. L. & Pol'y J.* 37 (2015), 543.

[4] Janine Berg, M. Furrer, E. Harmon, U. Rani, and M. S. Silberman. 2018. Digital labour platforms and the future of work: Towards decent work in the online world. *Geneva: International Labour Organization, September* 20 (2018).

[5] Michael S. Bernstein, Greg Little, Robert C. Miller, Björn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. 2015. Soylent: A word processor with a crowd inside. *Commun. ACM* 58, 8 (2015), 85–94. DOI:https://doi.org/10.1145/2791285

[6] M. R. Blais, L. Lachance, R. J. Vallerand, N. M. Briere, and A. S. Riddle. 1993. The work motivation inventory. *Revue Quebecoise de Psychologie* 14 (1993), 185–215.

[7] Kenneth A. Bollen. 1989. *Structural Equations with Latent Variables*. John Wiley & Sons.

[8] Robin Bordoli. 2018. Focused on the Future with a New Name. Retrieved from http://www.figure-eight.com/focused-future-new-name/.

[9] Dana Chandler and Adam Kapelner. 2013. Breaking monotony with meaning: Motivation in crowdsourcing markets. *J. Econ. Behav. Organiz.* 90 (2013), 123–133.

[10] Jonathan Chang, Jordan L. Boyd-Graber, Sean Gerrish, Chong Wang, and David M. Blei. 2009. Reading tea leaves: How humans interpret topic models. In *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems*, Yoshua Bengio, Dale Schuurmans, John D. Lafferty, Christopher K. I. Williams, and Aron Culotta (Eds.). Curran Associates, 288–296. Retrieved from http://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models.

[11] Emanuela Chemolli and Marylène Gagné. 2014. Evidence against the continuum structure underlying motivation measures derived from self-determination theory. *Psychol. Assess.* 26, 2 (2014), 575.

[12] Fang Fang Chen. 2007. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct. Equat. Model.* 14, 3 (2007), 464–504.

[13] Gordon W. Cheung and Roger B. Rensvold. 2000. Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *J. Cross-Cultur. Psychol.* 31, 2 (2000), 187–212.

[14] Gordon W. Cheung and Roger B. Rensvold. 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equat. Model.* 9, 2 (2002), 233–255.

[15] Cristiano Codagnone, Fabienne Abadie, and Federico Biagi. 2016. The future of work in the "sharing economy." Market efficiency and equitable opportunities or unfair precarisation? *Market Efficiency and Equitable Opportunities or Unfair Precarisation* (2016).

[16] European Commission. 2016. A European agenda for the collaborative economy. Retrieved from http://ec.europa.eu/DocsRoom/documents/16881.

[17] Lee J. Cronbach. 1951. Coefficient alpha and the internal structure of tests. *psychometrika* 16, 3 (1951), 297–334.

[18] Douglas P. Crowne and David Marlowe. 1960. A new scale of social desirability independent of psychopathology. *J. Consult. Psychol.* 24, 4 (1960), 349.

[19] María de los Ángeles Morata-Ramírez and Francisco Pablo Holgado-Tello. 2013. Construct validity of Likert scales through confirmatory factor analysis: A simulation study comparing different methods of estimation based on Pearson and polychoric correlations. *Int. J. Soc. Sci. Studies* 1, 1 (2013), p54–61.

[20] Edward L. Deci and Richard M. Ryan. 1980. The empirical exploration of intrinsic motivational processes. *Adv. Exper. Soc. Psychol.* 13 (1980), 39–80.

[21] Edward L. Deci and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior.* Springer US.

[22] Edward L. Deci and Richard M. Ryan. 2000. The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychol. Inquiry* 11, 4 (2000), 227–268.

[23] Edward L. Deci and Richard M. Ryan. 2002. *Handbook of Self-determination Research.* University Rochester Press.

[24] Edward L. Deci and Richard M. Ryan. 2002. Overview of self-determination theory: An organismic dialectical perspective. *Handbook Self-determ. Res.* (2002), 3–33.

[25] Robert F. DeVellis. 2016. *Scale Development: Theory and Applications.* Vol. 26. Sage publications.

[26] Christoph Dybowski and Sigrid Harendza. 2015. Validation of the physician teaching motivation questionnaire (PTMQ). *BMC Med. Edu.* 15, 1 (2015), 1.

[27] Alek Felstiner. 2011. Working the crowd: Employment and labor law in the crowdsourcing industry. *Berkeley J. Employ. Labor Law* (2011), 143–203.

[28] Sara J. Finney and Christine DiStefano. 2006. Non-normal and categorical data in structural equation modeling. *Structural equation modeling: A second course* (2006), 269–314.

[29] GESIS–Leibniz Institute for the Social Sciences. 2019. Retrieved from http://crowdworkers.info.

[30] Claes Fornell and David F. Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *J. Market. Res.* (1981), 39–50.

[31] James R. Foulds, Levi Boyles, Christopher DuBois, Padhraic Smyth, and Max Welling. 2013. Stochastic collapsed variational Bayesian inference for latent Dirichlet allocation. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'13),* Inderjit S. Dhillon, Yehuda Koren, Rayid Ghani, Ted E. Senator, Paul Bradley, Rajesh Parekh, Jingrui He, Robert L. Grossman, and Ramasamy Uthurusamy (Eds.). ACM, 446–454. DOI : https://doi.org/10.1145/2487575.2487697

[32] Robin D. Froman. 2001. Elements to consider in planning the use of factor analysis. *South. Online J. Nurs. Res.* 2, 5 (2001).

[33] Marylène Gagné and Edward L. Deci. 2005. Self-determination theory and work motivation. *J.f Organiz. Behav.* 26, 4 (2005), 331–362.

[34] Marylène Gagné, Jacques Forest, Marie-Hélène Gilbert, Caroline Aubé, Estelle Morin, and Angela Malorni. 2010. The motivation at work scale: Validation evidence in two languages. *Edu. Psychol. Measure.* 70, 4 (2010), 628–646.

[35] Marylène Gagné, Jacques Forest, Maarten Vansteenkiste, Laurence Crevier-Braud, Anja Van den Broeck, Ann Kristin Aspeli, Jenny Bellerose, Charles Benabou, Emanuela Chemolli, Stefan Tomas Güntert, et al. 2015. The multidimensional work motivation scale: Validation evidence in seven languages and nine countries. *Eur. J. Work Organiz. Psychol.* 24, 2 (2015), 178–196.

[36] Darren George and M. Mallery. 2003. Using SPSS for Windows step by step: A simple guide and reference. Allyn & Bacon, Boston, MA.

[37] Frédéric Guay, Alexandre J. S. Morin, David Litalien, Pierre Valois, and Robert J. Vallerand. 2015. Application of exploratory structural equation modeling to evaluate the academic motivation scale. *J. Exper. Edu.* 83, 1 (2015), 51–82.

[38] Neha Gupta, Andy Crabtree, Tom Rodden, David Martin, and Jacki O'Neill. 2014. Understanding Indian crowdworkers. In *Proceedings of the 17th Conference on Computer Supported Cooperative Work.*

[39] Neha Gupta, David B. Martin, Benjamin V. Hanrahan, and Jacki O'Neill. 2014. Turk-life in India. In *Proceedings of the 18th International Conference on Supporting Group Work,* Sean P. Goggins, Isa Jahnke, David W. McDonald, and Pernille Bjørn (Eds.). ACM, 1–11. DOI : https://doi.org/10.1145/2660398.2660403

[40] Joseph F. Hair, Joseph F. Jr., Barry J. Babin, Rolph E. Anderson, and William C. Black. 2018. *Multivariate Data Analysis (8th Ed.).* Cengage, Hampshire, United Kingdom.

[41] Kevin Hewison and Arne L. Kalleberg. 2013. Precarious work and flexibilization in South and Southeast Asia. *Amer. Behav. Sci.* 57, 4 (2013), 395–402.

[42] Daire Hooper, Joseph Coughlan, and Michael Mullen. 2008. Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods* 6 (2008), 53–60.

[43] John L. Horn and J. Jack McArdle. 1992. A practical and theoretical guide to measurement invariance in aging research. *Exper. Aging Res.* 18, 3 (1992), 117–144.

[44] John Joseph Horton and Lydia B. Chilton. 2010. The labor economics of paid crowdsourcing. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC'10)*, David C. Parkes, Chrysanthos Dellarocas, and Moshe Tennenholtz (Eds.). ACM, 209–218. DOI:https://doi.org/10.1145/1807342.1807376

[45] Mokter Hossain. 2012. Users' motivation to participate in online crowdsourcing platforms. In *Proceedings of the International Conference on Innovation Management and Technology Research (ICIMTR'12)*. IEEE, 310–315.

[46] Joshua L. Howard, Marylène Gagné, Alexandre J. S. Morin, and Jacques Forest. 2016. Using bifactor exploratory structural equation modeling to test for a continuum structure of motivation. *J. Manage.* (2016), 0149206316645653.

[47] Jeff Howe. 2006. Crowdsourcing: A Definition. Retrieved from http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html.

[48] Panagiotis G. Ipeirotis. 2010. Analyzing the Amazon mechanical turk marketplace. *ACM Crossroads* 17, 2 (2010), 16–21. DOI:https://doi.org/10.1145/1869086.1869094

[49] Panagiotis G. Ipeirotis. 2010. Demographics of mechanical turk. *CeDER Working Papers* (2010).

[50] Arne L. Kalleberg. 2009. Precarious work, insecure workers: Employment relations in transition. *Amer. Sociol. Rev.* 74, 1 (2009), 1–22.

[51] Nicolas Kaufmann, Thimo Schulze, and Daniel Veit. 2011. More than fun and money. Worker motivation in crowdsourcing—A study on mechanical turk. In *Proceedings of the 17th Americas Conference on Information Systems: A Renaissance of Information Technology for Sustainability and Global Competitiveness (AMCIS'11)*, Vallabh Sambamurthy and Mohan Tanniru (Eds.). Association for Information Systems. Retrieved from http://aisel.aisnet.org/amcis2011_submissions/340.

[52] Shashank Khanna, Aishwarya Ratan, James Davis, and William Thies. 2010. Evaluating and improving the usability of mechanical turk for low-income workers in India. In *Proceedings of the 1st ACM Annual Symposium on Computing for Development (ACM DEV'10)*, Andrew M. Dearden, Tapan S. Parikh, and Lakshminarayanan Subramanian (Eds.). ACM, 12. DOI:https://doi.org/10.1145/1926180.1926195

[53] Rex B. Kline. 2015. *Principles and Practice of Structural Equation Modeling*. Guilford Publications.

[54] Jon A. Krosnick. 1999. Survey research. *Ann. Rev. Psychol.* 50, 1 (1999), 537–567.

[55] Siou Chew Kuek, Cecilia Paradi-Guilford, Toks Fayomi, Saori Imaizumi, Panos Ipeirotis, Patricia Pina, Manpreet Singh, et al. 2015. *The Global Opportunity in Online Outsourcing*. Technical Report. The World Bank.

[56] Ihno A. Lee and Kristopher J. Preacher. 2013. Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Retrieved from http://quantpsy.org/corrtest/corrtest3.htm.

[57] Tak Yeon Lee, Casey Dugan, Werner Geyer, Tristan Ratchford, Jamie C. Rasmussen, N. Sadat Shami, and Stela Lupushor. 2013. Experiments on motivational feedback for crowdsourced workers. In *Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM'13)*. Emre Kiciman, Nicole B. Ellison, Bernie Hogan, Paul Resnick, and Ian Soboroff (Eds.). The AAAI Press. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6118.

[58] David Litalien, Frédéric Guay, and Alexandre J. S. Morin. 2015. Motivation for PhD studies: Scale development and validation. *Learn. Individ. Dif.* 41 (2015), 1–13.

[59] Irene Mandl, Maurizio Curtarelli, Sara Riso, Oscar Vargas, and Elias Gerogiannis. 2015. New forms of employment. *Eurofound Report*.

[60] Herbert W. Marsh, Kit-Tai Hau, and David Grayson. 2005. Goodness of fit in structural equation models. In *Multivariate applications book series. Contemporary Psychometrics: A Festschrift for Roderick P. McDonald*, A. Maydeu-Olivares and J. J. McArdle (Eds.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US.

[61] David B. Martin, Benjamin V. Hanrahan, Jacki O'Neill, and Neha Gupta. 2014. Being a turker. In *Proceedings of the Computer Supported Cooperative Work (CSCW'14)*, Susan R. Fussell, Wayne G. Lutters, Meredith Ringel Morris, and Madhu Reddy (Eds.). ACM, 224–235. DOI:https://doi.org/10.1145/2531602.2531663

[62] Roderick P. McDonald. 1999. *Test Theory: A Unified Treatment*. Lawrence Erlbaum Associates Publishers.

[63] Roger E. Millsap and Margarita Olivera-Aguilar. 2012. Investigating measurement invariance using confirmatory factor analysis. In *Handbook of Structural Equation Modeling*, R. Hoyle (Ed.). Guilford Press, New York.

[64] Babak Naderi. 2018. How to measure motivation? In *Motivation of Workers on Microtask Crowdsourcing Platforms*. Springer, 29–44.

[65] Babak Naderi, Ina Wechsung, Tim Polzehl, and Sebastian Möller. 2014. Development and validation of extrinsic motivation scale for crowdsourcing microtask platforms. In *Proceedings of the International ACM Workshop on Crowdsourcing for Multimedia (CrowdMM'14)*, Judith Redi and Mathias Lux (Eds.). ACM, 31–36. DOI:https://doi.org/10.1145/2660114.2660122

[66] Bloomberg News. 2016. Venezuela's Currency is Collapsing on the Black Market Again. Retrieved from http://www.bloomberg.com/news/articles/2016-11-01/venezuela-s-currency-is-collapsing-on-the-black-market-again.

[67] Ailsa G. Niven and David Markland. 2016. Using self-determination theory to understand motivation for walking: Instrument development and model testing using Bayesian structural equation modelling. *Psychol. Sport. Exercise* 23 (2016), 90–100.

[68] Kimberly A. Noels, Luc G. Pelletier, Richard Clément, and Robert J. Vallerand. 2000. Why are you learning a second language? Motivational orientations and self-determination theory. *Lang. Learn.* 50, 1 (2000), 57–85.

[69] Daniel M. Oppenheimer, Tom Meyvis, and Nicolas Davidenko. 2009. Instructional manipulation checks: Detecting satisficing to increase statistical power. *J. Exper. Soc. Psychol.* 45, 4 (2009), 867–872.

[70] Jason W. Osborne and Anna B. Costello. 2009. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Manage. Rev.* 12, 2 (2009), 131–146.

[71] Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. Running experiments on Amazon mechanical turk. *Judg. Decis. Mak.* 5, 5 (2010), 411–419.

[72] European Parliament. 2016. The situation of workers in the Collaborative Economy. Retrieved from http://www.europarl.europa.eu/RegData/etudes/IDAN/2016/587316/IPOL_IDA(2016)587316_EN.pdf.

[73] Lisa Posch, Arnim Bleier, Fabian Flöck, and Markus Strohmaier. 2018. Characterizing the global crowd workforce: A cross-country comparison of crowdworker demographics. arXiv preprint arXiv:1812.05948.

[74] Lisa Posch, Arnim Bleier, Philipp Schaer, and Markus Strohmaier. 2015. The polylingual labeled topic model. In *Proceedings of the 38th Annual German Conference on Advances in Artificial Intelligence (AI'15) (Lecture Notes in Computer Science)*, Steffen Hölldobler, Markus Krötzsch, Rafael Peñaloza, and Sebastian Rudolph (Eds.), Vol. 9324. Springer, 295–301. DOI : https://doi.org/10.1007/978-3-319-24489-1_26

[75] Lisa Posch, Arnim Bleier, and Markus Strohmaier. 2017. Measuring motivations of crowdworkers: The multidimensional crowdworker motivation scale (first version). *CoRR* http://arxiv.org/abs/1702.01661v1 (2017).

[76] Jakob Rogstadius, Vassilis Kostakos, Aniket Kittur, Boris Smus, Jim Laredo, and Maja Vukovic. 2011. An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2778.

[77] Joel Ross, Lilly Irani, M. Six Silberman, Andrew Zaldivar, and Bill Tomlinson. 2010. Who are the crowdworkers?: Shifting demographics in mechanical turk. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI'10)*, Elizabeth D. Mynatt, Don Schoner, Geraldine Fitzpatrick, Scott E. Hudson, W. Keith Edwards, and Tom Rodden (Eds.). ACM, 2863–2872. DOI : https://doi.org/10.1145/1753846.1753873

[78] Yves Rosseel. 2012. lavaan: An R package for structural equation modeling. *J. Stat. Software* 48, 2 (2012), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/.

[79] Richard M. Ryan and James P. Connell. 1989. Perceived locus of causality and internalization: Examining reasons for acting in two domains. *J. Personal. Soc. Psychol.* 57, 5 (1989), 749.

[80] Richard M. Ryan and Edward L. Deci. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemp. Edu. Psychol.* 25, 1 (2000), 54–67.

[81] Cristina Sarasua, Elena Simperl, and Natalya Fridman Noy. 2012. CrowdMap: Crowdsourcing ontology alignment with microtasks. In *Proceedings of the 11th International Semantic Web Conference on the Semantic Web (ISWC'12) (Lecture Notes in Computer Science)*, Philippe Cudré-Mauroux, Jeff Heflin, Evren Sirin, Tania Tudorache, Jérôme Euzenat, Manfred Hauswirth, Josiane Xavier Parreira, Jim Hendler, Guus Schreiber, Abraham Bernstein, and Eva Blomqvist (Eds.), Vol. 7649. Springer, 525–541. DOI : https://doi.org/10.1007/978-3-642-35176-1_33

[82] Albert Satorra and Peter M. Bentler. 2010. Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika* 75, 2 (2010), 243–248.

[83] semTools Contributors. 2016. *semTools: Useful Tools for Structural Equation Modeling*. Retrieved from http://CRAN.R-project.org/package=semTools.

[84] Kennon M. Sheldon and Andrew J. Elliot. 1998. Not all personal goals are personal: Comparing autonomous and controlled reasons for goals as predictors of effort and attainment. *Personal. Soc. Psychol. Bull.* 24, 5 (1998), 546–557.

[85] Alexander D. Stajkovic and Fred Luthans. 1997. A meta-analysis of the effects of organizational behavior modification on task performance, 1975–95. *Acad. Manage. J.* 40, 5 (1997), 1122–1149.

[86] James H. Steiger. 1980. Tests for comparing elements of a correlation matrix. *Psychol. Bull.* 87, 2 (1980), 245.

[87] Holger Steinmetz. 2013. Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences* 9, 1 (2013), 1–12.

[88] Maxime A. Tremblay, Céline M. Blanchard, Sara Taylor, Luc G. Pelletier, and Martin Villeneuve. 2009. Work extrinsic and intrinsic motivation scale: Its value for organizational psychology research. *Can. J. Behav. Sci.* 41, 4 (2009), 213.

[89] Amazon Mechanical Turk. 2018. Amazon Mechanical Turk: Worker Web Site FAQs. Retrieved from http://www.mturk.com/mturk/help?helpPage=worker#how_paid.

[90] Robert J. Vallerand, Luc G. Pelletier, Marc R. Blais, Nathalie M. Briere, Caroline Senecal, and Evelyne F. Vallieres. 1992. The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Edu. Psychol. Measure.* 52, 4 (1992), 1003–1017.

[91] Rens Van de Schoot, Peter Lugtig, and Joop Hox. 2012. A checklist for testing measurement invariance. *Eur. J. Dev. Psychol.* 9, 4 (2012), 486–492.

[92] David L. Vannette and Jon A. Krosnick. 2014. A comparison of survey satisficing and mindlessness. *Wiley Blackwell Handbook Mindful.* 1 (2014), 312.

[93] Roger L. Worthington and Tiffany A. Whittaker. 2006. Scale development research: A content analysis and recommendations for best practices. *Counsel. Psychol.* 34, 6 (2006), 806–838.