

Discussion Papers In Economics And Business

ワーク・エンゲイジメントスコアの国際比較における回答スタイルの影響

―日本とフランスを対象とした係留ビネット調査―

澤田 真央

Discussion Paper 26-15

November 2026

Graduate School of Economics
The University of Osaka, Toyonaka, Osaka 560-0043, JAPAN

ワーク・エンゲイジメントスコアの国際比較における回答スタイルの影響

―日本とフランスを対象とした係留ビネット調査―

澤田　真央[†]

要旨

日本人のワーク・エンゲイジメントスコアの国際的な低水準について，回答時のバイアスがどの程度寄与しているのかを，日仏の従業員を対象に係留ビネット法を用いた調査分析を行うことで，定量的に検討した。その結果，日本の平均スコアは仏に対し 1.27 点低く，その差の約 3 割を回答時のバイアスが説明した。これは主に仏側の上位選択肢を選びやすい傾向によるものであった。一方，約 7 割の差は補正後も残留し，日本の WE 低水準は主として実質的な差を反映することが示唆される。

JEL 分類：O15 J24 J28

キーワード：ワーク・エンゲイジメント, UWES 尺度, レスポンス・スタイル, 係留ビネット法, 多次元項目反応理論

[†] 大阪大学大学院経済学研究科　経営学系専攻　博士後期課程　mail: u338546k@ecs.osaka-u.ac.jp

1. はじめに

本研究の目的は，日本人労働者のワーク・エンゲイジメント（Work Engagement：以下 WE）の国際的な低水準に対して，回答時のバイアスがどの程度寄与しているのかを，日本人およびフランス人従業員を対象に調査を行い， 係留ビネット法（Anchoring Vignettes）と多次元 IRT(Item Response Theory, 以下 IRT)モデルを用いて定量的に検討することである。

WE は「仕事に関連するポジティブで充実した心理状態」と定義され，従業員の健康や生産性，組織成果と関連する重要な概念として近年注目されている。その一方で WE の国際比較研究では，日本人の WE スコアが他国と比較して著しく低いことが報告されている（Shimazu et al., 2010）。その要因として，集団の調和を重視する文化的背景から，ポジティブな感情や態度の表出を抑制する傾向があることを指摘しており（Iwata et al., 1995），このような回答時に生じるバイアスが WE スコアの低水準に影響している可能性が示唆されている。

代表的な回答バイアスであるレスポンス・スタイル（Response Style：以下 RS）は心理尺度の信頼性や妥当性を脅かす要因とされている。しかし，RS が WE スコアをどの程度歪めているのか，また国際比較で観察される日本人の低スコアをどの程度説明しうるのかについては十分に検討されていない。

そこで本研究では，日本とフランスの就業者を対象に係留ビネット法を用いた WE スコアの調査を実施し，WE 国際比較における RS の影響を検証する。

2. 理論的背景と問題意識

2.1 ワーク・エンゲイジメントと国際比較研究

WE は，特定の出来事や対象に一時的に生じる感情ではなく，仕事に向けられた持続的かつ全般的な感情と認知であり，活力（vigor），熱意（dedication），没頭（absorption）という 3 つの側面によって特徴づけられる（Schaufeli et al., 2002; Schaufeli & Bakker, 2004）。本研究においても，先行研究と同様にこの定義に従うものとする。従業員の WE が高まると職務満足度や組織コミットメントが高まり，仕事のパフォーマンス向上につながるため経営学的に重要な指標である。

WE を測定する代表的な尺度である Utrecht Work Engagement Scale（以下，UWES）は, 3 つの下位次元から構成され, 17 項目で活力，熱意，没頭を測定する(Schaufeli et al., 2002)。これまでに, 32 カ国で標準化または使用されており，日本語版を含む，いずれの言語においても，良好な信頼性および妥当性が確認されている(島津, 2025; Shimazu et al., 2008)。

WE スコアを国際比較した研究では，上述のように日本の WE スコアが著しく低い値となっ

ていることが示されている。Shimazu et al. (2010) は, UWES の短縮版（9 項目版)を用いて，日本を含む 16 カ国において WE スコア比較しており，そのなかで日本のスコア平均は 2 点台後半と最下位であったのに対し，他の 15 カ国は 3 点台後半を上回っており，日本の次に低い値であるスペインでは 3 点台後半,最上位のフランスでは 4 点台後半となっていた。もし日本人従業員の WE が実際に他国と比較して低いのであれば，その背景にある組織的要因や仕事の資源の不足などといった職場環境の他国との差異を具体的に検討することが必要である。さらには，日本の従業員の WE を高めていくための，組織による適切な介入や施策を検討することが重要な社会課題である。

一方で，先行研究ではこの要因として文化的なバイアスの影響を指摘している(Shimazu et al., 2010)。日本では集団の調和を重視する文化的背景から，ポジティブな感情や態度の表出を抑制する傾向があることを指摘しており（Iwata et al., 1995），こうした文化的傾向が WE スコアの低水準の一因となっている可能性を示唆している。実務的には，多くの企業が従業員の WE 向上を目的として 1on1 面談や研修プログラムといった人的資源への施策を導入しているが，その効果は WE が実態として低いことを前提として成立する。もし低スコアの一因が実態を反映しない回答バイアスにあるとすれば，問題の所在を誤って判断し施策を実施することになり，組織にとっては資源配分が非効率になるだけでなく，従業員にとっては介入の効果を十分に得られないことになる。また，スコアの差が測定上の問題によって生じているのであれば，これまでの国際比較研究の解釈や，WE 向上に向けた議論の前提そのものを再検討する必要がある。いずれの場合においても，測定上のバイアスの影響を切り分けることは，日本の組織特性と WE との関係を実証的に論じるための前提条件として不可欠である。

2.2 回答バイアスおよび本研究の仮説

回答バイアスとは，本来測定したい内容とは異なる，項目への回答にあたっての系統的な傾向（Paulhus,1991)である。回答バイアスには，質問内容と無関係に見られる個人に特徴的な回答傾向に注目しているバイアスと個々の回答者に依存せず項目内容に依存するものがあり，RS は前者に分類される（Paulhus,1991; 北條・岡田,2018)。RS は質問の内容に関係なく，特定の選択肢を選択する傾向と定義される(Johnson et al., 2011; Yang et al., 2010)。大きく，黙従反応傾向（acquiescence response style），極端反応傾向（extreme response style）,中間反応傾向（midpoint response style）などの分類がある (Yang et al., 2010)。

RS は, 平均値や相関係数, 回帰係数などの推定値を歪める可能性があるため, 調査結果の妥当性に対する脅威となる（Van Vaerenbergh & Thomas, 2013)。さらに, RS は文化的・社会的要因と関連しており, 国際調査においては, 比較可能性を阻む要因となる（Kankaraš &

Moors, 2011)。また，項目内容に依存するようなバイアスについても国によって内容の解釈に差が生じると国際比較が困難となる要因となりうる。

なかでも，日本人の RS に関しては，極端な回答を避け，中間的な回答を好む傾向にあることなどが複数の先行研究において報告されている(Chen et al., 1995; 米村・松田, 1991; 田崎・申, 2017)。また，26 カ国で国際比較した研究では日本が最も中間回答を好み，黙従傾向が低いことが示されている(Harzing, 2006)。フランスは日本とは対照的に 26 カ国中最も中間回答を好まず，肯定的な選択肢を選ぶ反応傾向が国際平均よりも高いことが明らかになっている（Harzing, 2006）。

このように文化的背景が異なる国の回答者には異なる RS の特徴が見られ，WE スコアの水準にも影響を与えうることが予想される。そこで本研究では，日本との比較対象として WE スコアの水準および RS の特徴が大きく異なるフランスを対象として，採用した。バイアスを補正する方法として係留ビネット法を用いた日仏のデータをもとに以下の仮説を検討する。

H1：補正前のフランスの WE スコアは日本の WE スコアよりも高い。

H2：RS 補正前後で日本人およびフランス人の WE スコアは有意に変化する。

H3：RS 補正後の日仏間の WE スコア差は補正前と比較して縮小する。

3. 方法

3.1 調査対象と実施手順

調査実施に際し，研究目的，個人情報の取扱い，参加の任意性等を説明した告知文を提示し，同意した回答者のみが質問票へ進める方式を採用した。氏名や住所等の個人を特定できる情報は収集しておらず，データは匿名化された状態で研究者に提供された。

調査は，アイブリッジ株式会社が運営する調査サービス Freeasy の登録モニターを対象に，2026 年 6 月にオンラインで実施した。サンプルは各国で就業する 20 歳以上の正規雇用の会社員を対象とし，年齢階級別・性別の労働力人口比に応じて割付を行った。回答は日仏それぞれ 501 名から得られた。

回答の質を確保するため，次の 2 つの基準のいずれかに該当する回答者を不誠実回答者として分析から除外した。第 1 に，後述する指示型項目 3 問のうち 2 問以上で指示と異なる選択肢を選んだ者である。第 2 に，WE のいずれかの次元に対応する質問群において，指示型

項目を除く 18 項目すべてに同一の選択肢を選び続けた者（いわゆるストレートライン回答）である。その結果，有効回答は日本 304 名（60.7%），フランス 165 名（32.9%）となった。フランスの有効回答率が低いのは，主として指示型項目での不誠実回答の排除によるものであった。除外基準は両国で同一とした。

3.2 調査設計

調査票は，WE 尺度，ビネット項目，指示型項目の 3 つを中心に構成し，回答者の属性についても収集した。回答者には，WE の次元ごとに，まず自身の WE について回答を求めた後，当該次元に対応する架空人物の説明（ビネット）の評価を求めた。

（1）基本属性： 年齢，性別，婚姻，居住地，業種，子供の有無，世帯年収，居住形態を収集した。

（2）WE 尺度： WE の測定には UWES の 9 項目版（UWES-9; Schaufeli et al., 2006）を用いた。UWES-9 は活力，熱意，没頭の 3 次元各 3 項目の計 9 項目からなり，回答は「0＝全くない」から「6＝いつも感じる」までの 7 件法で求めた。得点が高いほど WE が高いことを示す。UWES-9 は多くの国・地域で使用されており，日本語版を含めて十分な信頼性および妥当性が確認されている（Schaufeli et al., 2006; Shimazu et al., 2008）。

（3）ビネット項目： RS を推定するため，King et al.（2004）の係留ビネット法に基づき，WE の 3 次元それぞれについて，高水準から低水準までの状態を表す 5 名の架空人物（A〜E）を描写した短文（ビネット）を作成した。例えば活力の次元では，「仕事への取りかかりの速さ」「仕事を継続する力」「困難な作業への粘り強さ」といった特徴の水準を段階的に変えて人物を描写した。回答者には，各人物が仕事に関してどの程度その質問文の状態に当てはまるかを，自己評定と同一の 7 件法（0〜6）で，1 人物あたり 3 項目，1 次元あたり計 15 項目で評定するよう求めた。筆者らの先行研究では 1 次元あたり 2 人物と少ないビネット項目数にとどまり RS の推定精度に課題があったため，本研究ではビネットを 5 名に増やした。なお，フランス語版のビネットは日本語版を翻訳して作成し，逆翻訳により原文との意味内容の整合性を確認した。

（4）指示型項目： 不誠実回答を検出するため，「この質問では選択肢 2 を選んでください」のように特定の選択肢の選択を明示的に指示する項目を，各次元の項目群に 1 問ずつ（計 3 問）設けた。

なお，調査票には個人レベルの文化的価値観を測定する項目も含めたが，紙幅の制約から，本報告では分析対象を RS と WE スコアの国際比較に限定する。文化的価値観と RS の関連については稿を改めて報告する予定である。

3.3 分析方法

分析には，Bock（1972）の名義反応モデルを基礎として係留ビネット法と組み合わせた多次元項目反応理論モデル（Bolt et al., 2014; 北條・岡田, 2017, 2018）を用いた。このモデルでは，回答者rが項目iで選択肢k（0〜6 の 7 段階）を回答Uとして選ぶ確率を次式で表す。

$$P(U_{ri} = k) \propto \exp(a_k\theta_r + s_{rk} + c_{ik})$$

ここで，θ_rは回答者rの WE の潜在的な水準（以下，潜在特性）である。a_kは選択肢kに割り当てた固定間隔の重み$(-3, -2, -1, 0, 1, 2, 3)$であり，潜在特性が高い回答者ほど上位の選択肢を選びやすくなる関係を表す。c_{ik}は項目iにおいて選択肢kがもともと持つ選ばれやすさ（項目の内容に依存する切片）である。s_{rk}は回答者rが項目の内容と無関係に選択肢kを好む程度であり，これが RS を表す。例えば中間の選択肢を好む回答者は中央のsが正に，両端の s が負に推定される。確率は，各選択肢の exp(・)の値を全選択肢の合計で割ることにより定まる。

また，推定の際はビネット項目のa_kをすべて 0 に固定する。ビネットに描かれた人物像は全回答者に共通であると仮定するため，その評定が回答者間で異なるとすれば，それは回答者の選択肢の使い方の違い，すなわちs（RS）に由来すると考えることができる。これにより，自己評定から RS の影響sを分離し潜在特性θを取り出すことができる（King et al., 2004; Bolt et al., 2014）。

本研究では，このモデルに次の 3 点の拡張を加えた。第 1 に，UWES-9 の 9 項目が 3 次元に分かれていることに対応して，潜在特性を，全項目に共通する WE 全体の因子（一般因子）と各次元に固有な因子とに分けて推定する双因子（bifactor）構造を採用した。国際比較の対象は一般因子を用いる。第 2 に，同じ項目でも国によって内容の受け取られ方が異なる項目機能差異（Differential Item Functioning：以下 DIF）を考慮し，項目切片c_{ik}の日仏差を表すパラメータを推定した。第 3 に，潜在特性と RS の平均およびばらつきは国ごとに推定し，特に潜在特性についてはフランスとの比較のため日本の平均を基準とする制約を課した。

パラメータの推定にはベイズ推定を用い，マルコフ連鎖モンテカルロ法（Markov chain Monte Carlo methods：以下 MCMC）により各パラメータの事後分布を求めた。独立な 4 つの Chain から各々1000 回のウォームアップ期間を経たのちに 2000 回ずつのサンプリングを行い，収束を標準的な診断指標（Vehtari et al., 2021）により確認した（全パラメータで収束指標$\hat{R} < 1.01$，発散遷移なし）。以下で報告する推定値は事後分布の平均であり，［ ］内の表記は 95%信用区間を示す。

RS および θ の双因子構造の考慮による影響を確かめるため，(i) θが単次元で RS を考慮しないモデル，(ii) θが単次元で RS を考慮したモデル，(iii)（ii）に双因子構造を加えたモデル，(iv)（iii）に項目切片の日仏差を加えたモデル，の 4 つを推定し，未知データへの予測力を表す指標（PSIS-LOO; Vehtari et al., 2017）を用いてモデルの比較を行った。

WE スコアの RS 補正については，最良モデルの推定結果において RS の項s_{rk}を 0 とした場合に期待される自己評定 9 項目の平均値（0〜6 点）として，回答者ごとに算出し，国ごとに平均を求めた。同様に，項目切片の日仏差のみを取り除いた場合，RS と項目切片の日仏差の両方を取り除いた場合のスコアも算出した。これにより，観測された日仏差を，(a) RS に由来する部分，(b) 項目内容の受け取り方の国ごとの差に由来する部分，(c) 両者を取り除いてもなお残る部分（潜在特性の差を反映すると解釈される部分），に切り分けて評価した。これらの量はすべて MCMC の事後標本ごとに計算しているため，不確実性は信用区間として評価した。

4. 結果

4.1 記述統計（仮説 H1 の検討）

有効回答者の国ごとの WE スコア（自己評定 9 項目の平均，範囲：0〜6 点）は，日本で平均 2.53（SD：1.48），フランスで平均 3.80（SD：1.19）であり，9 項目の内的整合性（Cronbach のα）は日本.96，フランス.93 と高い水準であった。日仏のスコア差は 1.27 点（効果量 Cohen のdでは 0.92）であり，補正前のフランスのスコアは日本を明確に上回った。したがって仮説 H1 は支持され，スコア差の方向および日本のスコア水準は，フランスが最上位・日本が最下位であった Shimazu et al.（2010）の 16 カ国比較と一致した。

4.2 モデル比較：RS は回答に影響しているか

4 つのモデルを予測の観点で比較したものを表 1 に示す。RS を含むモデル（ii）は RS を含まないモデル（i）に比べ，予測力が大幅に高かった（差＝＋10,681，差の標準誤差 131）。双因子構造の追加（＋216，標準誤差 28），項目切片の日仏差の追加（＋58，標準誤差 11）も，それぞれ予測力を改善した。これらの差はいずれも標準誤差を大きく上回っており，(a) RS が回答に強く影響していること，(b) 項目内容の受け取り方にも小さいながら確かな国差が存在すること，が示された。次節以降の結果はすべて最良モデル（iv）に基づく。

表 1　モデル比較（PSIS-LOO の計算結果と，最良モデルとのモデル間差）

モデル	予測力（$elpd_{\mathrm{LOO}}$，高いほど良い）	最良モデル（iv）との差（±標準誤差）
（i）RS なし・単次元	−42,160.7（±108.4）	−10,955.1（±131.4）
（ii）RS あり・単次元	−31,479.5（±138.4）	−273.9（±27.8）
（iii）（ii）＋双因子構造	−31,264.0（±140.2）	−58.4（±11.2）
（iv）（iii）＋項目切片の日仏差	−31,205.6（±140.2）	0（最良）

4.3 RS 補正前後の WE スコアの変化（H2・H3 の検討）

RS 補正（モデル上で RS の項を取り除く操作）による WE スコアの国ごとの平均の変化は，日本では−0.03 点［−0.09, 0.03］と 95%信用区間が 0 を含み，明確な変化は認められなかった。一方フランスでは−0.39 点［−0.47, −0.31］と明確な低下が認められた。したがって，補正前後でスコアが変化するという H2 は，フランスについてのみ支持され，全体としては部分的な支持にとどまった。補正によるスコア分布の変化を図 1 に示す。日本では補正前後の分布がほぼ重なるのに対し，フランスでは分布全体が低い方向へ移動しており，補正の効果がほぼフランス側に集中していることが分布の形からも確認できる。回答者ごとにみた RS の影響（補正前スコア − 補正後スコア）を調べたところ，その 95%信用区間が 0 を含まない，すなわち RS の影響が明確な回答者は，日本で 51%，フランスで 59%にのぼった。日本では押し上げと押し下げが個人間でほぼ相殺するのに対し，フランスでは押し上げ方向に偏っていた。

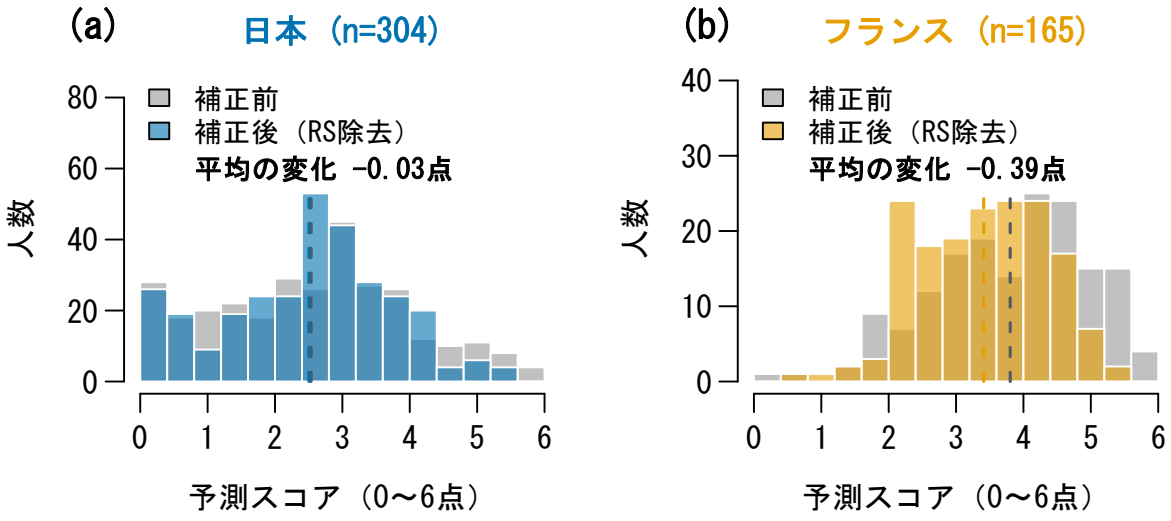

図 1　反応スタイル補正による回答者ごとの予測スコア分布の変化。
（a）日本，（b）フランス。破線＝それぞれの平均。

日仏のスコア差については図 2 に整理した。補正前の差 1.27 点は RS 補正後に 0.91 点［0.83, 0.99］へと縮小した。縮小幅は 0.36 点［0.30, 0.42］であり，補正前の差の 28.5%に相当する。したがって H3 は支持された。さらに，項目切片の日仏差（項目内容の受け取り方の差）を取り除いても日仏差はほとんど変わらず（0.90 点［0.81, 1.00］），その寄与は 0.002 点［−0.034, 0.037］と無視できる水準であった。一方で，RS および項目切片の日仏差の両者を補正してもなお残った，元のスコア差の約 7 割の部分は（効果量dは 0.75［0.67, 0.82］），本モデルの仮定の下で，潜在特性すなわち WE の実質的な水準差を反映するものであると予想される。

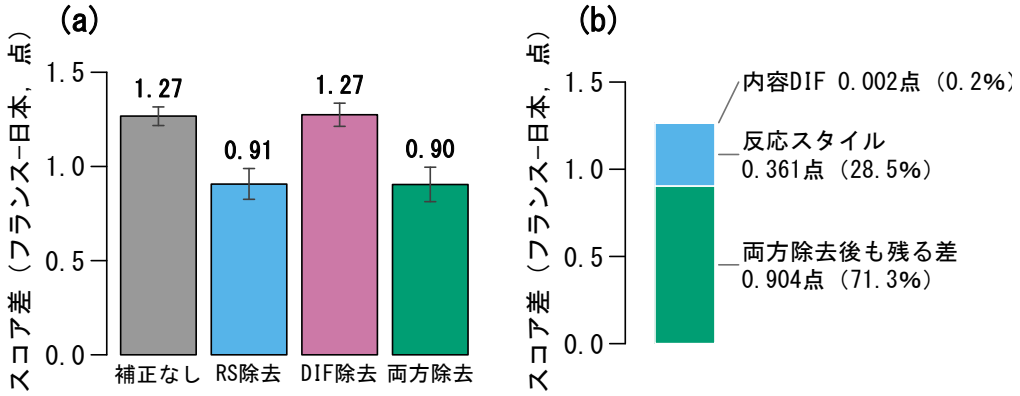

図 2　補正条件ごとの日仏の WE スコア差（a）と，観測されたスコア差 1.27 点の内訳（b）。エラーバーは 95%信用区間。DIF＝項目機能差異。

4.4 潜在特性の分布

RS と項目内容の国差を取り除いた後の潜在特性（一般因子）について，日本の平均を基準（0）としたフランスの平均の差は 1.20［0.88, 1.53］であり，明確に正であった（この値はモデル上の潜在尺度での差であり，0〜6 点のスコア差 0.90 点とは尺度が異なる）。特筆すべきは，日本にのみ著しく低い水準の集団が特徴的に多く観察された。これは平均の比較だけでは捉えられない分布の形の違いであり，次節で述べる測定精度の限界に留意しつつも，極端に WE が低い層の存在として今後の検討に値する。

4.5　補正の前提と測定精度の確認

係留ビネット法は，(a) ビネットが全回答者に同様に受け取られること（ビネット等価性），(b) 回答者が自己評定とビネット評定とで同じ選択肢の使い方をすること（反応一貫性）を前提とする（King et al., 2004）。本データでは，5 人物の平均評定の順序は両国とも想

定どおりであり，国とビネットの交互作用も小さく（分散でみると 0.4%以下の説明），ビネット等価性は日仏間では概ね満たされていた。ただし，ビネット評定には年齢層による差が特に日本で認められた。一方，反応一貫性の観点では，自己評定とビネット評定から推定される RS の一致は中程度（相関 .38〜.57）にとどまった。

5. 考察

5.1 結果の要約と解釈

本研究の主な結果は次の 3 点である。第 1 に，モデル比較の結果，RS は回答に確かに影響しており，補正の結果，日仏の WE スコア差の約 3 割を説明した。第 2 に，補正後に残留した約 7 割のスコア差は RS と項目内容の受け取り方の国差を取り除いてもなお残り，効果量にして Cohen のdが 0.75 という大きな差であった。第 3 に，補正によるスコアの変化は専らフランス側で生じ（−0.39 点），日本側のスコアはほとんど変化しなかった（−0.03 点）。

特筆すべきは第 3 の点である。日本人の WE スコアの低さについては，謙遜や中庸を好む文化的傾向による過小報告が一因である可能性がしばしば指摘されてきた。しかし本研究の結果が示すのは，日仏差のうち RS に由来する部分は，日本人の控えめな回答によるものではなく，フランス人回答者が上位の選択肢を選びやすい傾向があることに起因するということであり，先行研究の主張は日本ではなく海外の回答傾向にこそ目を向けられるものであることを示している。

そのうえで，補正後もなお残る 0.90 点の差は，日本人就業者の WE の低さの大部分が，測定上のバイアスではなく実質的な水準差を反映している可能性が高いことを示す。

5.2 学術的・実践的示唆

学術的には，本研究は，筆者らの知る限り，係留ビネット法と多次元 IRT モデルを WE の国際比較に適用し，観測スコア差を RS 由来・項目内容由来・残差に分解して示した最初の試みである。特に，複数の国の WE 水準の比較に，RS がどの程度スコア水準に影響を与えるのかについて定量的な知見を提供したという点と，項目切片の国間差の存在が支持されながらも，尺度全体の平均差にはほとんど影響しないという知見は，WE 国際比較における測定不変性の議論に対して有用な知見を示すものである。

実践的には，第 1 に，日本の従業員の WE が回答バイアスにより本当の値より低くスコアに表れると考える根拠は得られなかった。したがって，低水準の背景にある職務資源や職場環境の要因を特定し，改善に取り組むことが依然として重要であることが裏づけられたとい

える。第 2 に，グローバル企業が国をまたいでエンゲイジメント・サーベイの結果を比較する際には，RS の影響，とりわけ高スコア国側のレスポンス・スタイルを考慮しなければ，拠点間の差を過大に見積もるおそれがある。第 3 に，個人レベルでは RS の影響が明確な回答者が約半数（日本 51%・フランス 59%）存在したことから，通常の WE スコアの調査結果を個人単位で人事的な判断に用いる場合には，集団平均の比較以上に回答傾向の影響に注意が必要である。

5.3 限界と今後の課題

本研究の限界として，以下の点が挙げられる。第 1 に，反応一貫性の成立が中程度にとどまった点である。このため，RS の寄与 28.5%という数値は，この前提に依存した推定値であることを念頭に解釈する必要がある。ただし，ビネット等価性は日仏間では概ね満たされており，日仏差の大半が補正後も残るという結論の方向は頑健と考えられる。今後は，行動の記述に徹したビネット文面の改良などにより一貫性を高めることが課題である。第 2 に，ビネットの評定には年齢層による差が認められたため，年齢層間の比較に本ビネットを用いることには慎重になる必要がある。第 3 に，調査はオンラインパネルを対象としており，割付により人口構成を近似したものの，その国のサンプルとしての代表として考えることには限界がある。また，フランスの有効回答率が不誠実回答の除外により低く，有効サンプルの偏りが生じている可能性は否定できない。この点は，フランスの平均スコアが Shimazu et al.（2010）の 4 点台後半を下回った一因の可能性もある。第 4 に，本報告では同時に測定した文化的価値観の分析を扱っておらず，RS の背景にある文化的要因の検討は今後の課題であり，また，WE の水準差の原因（職務資源等）の特定には別途の調査および検討が必要である。

【文献リスト】

北條大樹・岡田謙介（2017）.「評定尺度における反応傾向を考慮した係留寸描データのベイズ的項目反応モデル」データ分析の理論と応用，6(1)，113-125.

北條大樹・岡田謙介（2018）.「係留ビネット法による反応スタイルの分類」行動計量学，45(1)，13-25.

島津明人. (2025). ワーク・エンゲイジメント: 研究と実践における最近 10 年の動向. 産業医学レビュー, 37(3), 121-142.

田崎勝也・申知元（2017）.「日本人の回答バイアス――レスポンス・スタイルの種別間・文化間比較――」心理学研究，88(1)，32-42.

米村恵子・松田義幸 (1991).「社会調査に観る日本人の曖昧・中間回答に関わる研究」日本社会心理学会第 32 回大会発表論文集, 66–69.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29-51.

Bolt, D. M., Lu, Y., & Kim, J.-S. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods*, 19(4), 528-541.

Chen, C., Lee, S.-Y., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6(3), 170-175.

Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, 6(2), 243-266.

Iwata, N., Roberts, C. R., & Kawakami, N. (1995). Japan-U.S. comparison of responses to depression scale items among adult workers. *Psychiatry Research*, 58(3), 237-245.

Johnson, T. P., Shavitt, S., & Holbrook, A. L. (2011) . Survey response styles across cultures. In D. Matsumoto & F. J. R. van de Vijver (Eds), *Cross-cultural research methods in psychology*. New York: Cambridge University Press. pp. 130- 175.

Kankaraš, M., & Moors, G. (2011). Measurement equivalence and extreme response bias in the comparison of attitudes across Europe: A multigroup latent-class factor approach. *Methodology*, 7(2), 68-80.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1),

191-207.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). Academic Press.

Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, 3(1), 71-92.

Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*, 25(3), 293-315.

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, 66(4), 701-716.

Shimazu, A., Schaufeli, W. B., Kosugi, S., Suzuki, A., Nashiwa, H., Kato, A., Sakamoto, M., Irimajiri, H., Amano, S., Hirohata, K., & Goto, R. (2008). Work engagement in Japan: Validation of the Japanese version of the Utrecht Work Engagement Scale. *Applied Psychology: An International Review*, 57(3), 510-523.

Shimazu, A., Schaufeli, W. B., Miyanaka, D., & Iwata, N. (2010). Why Japanese workers show low work engagement: An item response theory analysis of the Utrecht Work Engagement Scale. *BioPsychoSocial Medicine*, 4, Article 17.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2), 195-217.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413-1432.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\hat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, 16(2), 667-718.

Yang, Y., Harkness, J. A., Chin, T.-Y., & Villar, A. (2010). Response styles and culture. In J. A. Harkness et al. (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 203-223). John Wiley & Sons.

The Impact of Response Styles on International Comparisons of Work Engagement Scores

—An Anchoring Vignette Study of Japan and France—

Mao SAWADA[†]

Abstract

This study quantitatively examines the extent to which response biases contribute to the internationally low work engagement (WE) scores observed among Japanese employees. Using the anchoring vignette method, we conducted a survey of employees in Japan and France to assess the impact of response styles on cross-national differences in WE scores. The results showed that the average WE score in Japan was 1.27 points lower than that in France, with response biases accounting for approximately 30% of this difference. This effect was primarily attributable to the tendency of French respondents to select higher response categories. However, approximately 70% of the difference remained after adjustment, suggesting that Japan's relatively low WE scores largely reflect substantive differences in work engagement rather than response biases.

JEL code：O15 J24 J28

Keywords：Work Engagement, Utrecht Work Engagement Scale, Response Style, Anchoring Vignette Method, Multidimensional IRT Model

[†] Graduate Student, Graduate School of Economics, The University of Osaka
mail:u338546k@ecs.osaka-u.ac.jp