

個人特性が経験サンプリング法における回答数の時間変化に及ぼす影響：StudentLife データセットを用いた実証分析

富永 登夢^{1,a)} 山本 修平^{1,b)} 倉島 健^{1,c)} 戸田 浩之^{1,†1,d)}

受付日 2022年5月5日, 採録日 2022年11月8日

概要：経験サンプリング法 (Experience Sampling Method, 以下 ESM) の質問に対する回答率には個人差があり, 特定の属性を持つ被験者からの回答が偏って集められてしまうという問題が指摘されている. ESM の質問に対する回答の仕方と被験者の属性の関係を分析する従来研究の多くは ESM 実施期間全体の回答数を調査している. 一方で, 回答数の時間的な変化についてはほとんど明らかにされていない. そこで本研究では, StudentLife のデータセットを用いて, 被験者の性格 (神経症傾向, 外向性, 開放性, 調和性, 誠実性) および心理状態 (うつ傾向, 主観的ストレスレベル, 孤独感, 人生満足度) と ESM に対する回答数の時間変化との関係を調べた. 回帰モデルと潜在成長曲線モデルを用いた分析の結果, ストレスレベルが高いと申告した被験者ほど ESM に対する総回答数が多かったが, 時間の経過とともに強く減少する傾向が見られ, 特に中間試験のようなストレスのかかる出来事の後に急に減少することが分かった. また, 内向的な (外向性の低い) 被験者は, 時間の経過に対する回答数の減少傾向は弱いものの, 学校祭のような社会的なイベントの後に回答する頻度が下がることが分かった. これらの結果をもとに, ESM で収集される回答データの個人特性や時間に対する不均一性について論じ, 将来的な ESM 研究における推奨事項を示した.

キーワード：経験サンプリング法, 人格, 心理, 重回帰分析, 潜在成長曲線モデル

Understanding How Personal Characteristics Affect Temporal Response Patterns in Experience Sampling Method: A Case Study of StudentLife Dataset

TOMU TOMINAGA^{1,a)} SHUHEI YAMAMOTO^{1,b)} TAKESHI KURASHIMA^{1,c)} HIROYUKI TODA^{1,†1,d)}

Received: May 5, 2022, Accepted: November 8, 2022

Abstract: Response rates to Experience Sampling Method (ESM) questions vary among individual subjects, which inevitably results in a disproportionate distribution of the ESM data volume to personal characteristics. Previous research has explored the relationship between response rates and personal characteristics; however, few studies focused on temporal patterns of response rates to ESM questions. In this study, using StudentLife dataset, we examined how personality traits (neuroticism, extraversion, openness, agreeableness, conscientiousness) and psychological status (depression, stress, loneliness, life satisfaction) relate to temporal response patterns in ESM. Primary results of our analyses using regression models and latent growth curve models indicated that the subjects with high stress levels had a higher total number of ESM responses, but their responses tended to decrease strongly throughout the experiment, especially after a stressful event such as the midterm examination. We also found that the subjects with low extraversion had a weak tendency to decrease the number of ESM responses throughout the experiment, but their responses decreased strongly after a social event such as the school festival. Based on our results, we discussed how the data collected with ESM studies are biased not only in terms of personal characteristics but also with timings.

Keywords: experience sampling method, personality, regression model, latent growth curve model

1. はじめに

経験サンプリング法 (Experience Sampling Method, 以下 ESM) とは, 被験者にその時その場の状況における思考, 感情, 行動, 経験に関する質問を尋ね, リアルタイムに回答させるというプロセスを継続的に行う実験手法である. 実験室で実施されるような伝統的な質問紙調査とは異なり, ESM は日常生活を過ごす被験者を対象とするため, 生態学的妥当性を高め, 想起バイアスを低減させながら, ある個人における主観的な状態の継時的な変化をとらえることができる. ESM は, このような主観的な状態の時間的な被験者内変動を表すデータを集める手法として認められており, 医学的な検証 [30], [35] や気分の予測 [22], [26] など, 多くの研究で利用されている.

信頼性の高い ESM データセットを構築するためには, その規模だけでなく, データの質の均一性が重要である. すなわち, 様々な状況における様々な被験者からの回答が一様に含まれていることが理想的である. しかし, ESM 研究は日常生活を過ごす被験者を継続的に追跡するという性質上, 人や状況に関する条件を動的に統制することが困難であるため, 実際には特定の条件にあてはまる回答データが偏って回収されやすく, ESM データセットは不均一になることが多い. たとえば, 精神的な問題をかかえる被験者 [19] や年齢が若い被験者 [20], [36] は ESM における回答率が低いという結果が従来研究で報告されている. これはすなわち, ESM への回答は精神的に健全で年齢が高いという条件にあてはまる被験者から偏って収集されやすいことを示唆している. このような ESM データの偏りを明らかにできれば, ESM データから得られる知見や技術の限界や一般性を正当に評価し議論できる. したがって, ESM データが偏る条件を記述することは, ESM 研究の根幹に関わる重要な課題であるといえる.

この課題に取り組む従来研究 [19], [20], [36] において, 被験者の個人的な性質と ESM 回答率の関係は調査されてきたが, 個人的な性質が ESM 回答率の時間変化に及ぼす影響については明らかにされていない. 我々は, たとえばストレスをかかえる被験者の実験期間全体にわたる回答数は多かったとしても, そのような被験者が緊張や不安がもたらされるような出来事を経験するときにはそれに対処することに集中するため ESM の質問に回答する時間を持たな

くなり, そのようなタイミングでその被験者からの回答数が減る, といったことがあるのではないかと想定している. このような事実が個人的な性質と回答率の時間変化の関係性の検証によって明らかになれば, その検証結果は個人的な性質だけでなく時間という観点から ESM のデータに偏りが生じる条件を記述することに貢献する. また, 特定の性質を持つ被験者の回答数が減るタイミングを事前に把握し, 対処するための有益な知見となることが期待される.

そこで本研究は, ESM データにおける被験者の個人的な性質や時間による偏りを明らかにすることを目的とし, “被験者の性格や心理状態が ESM 質問回答数の時間変化にどのような影響を及ぼすか?” という研究課題に取り組む. 具体的には, StudentLife データセット [39] を利用した実証分析を行い, 得られた知見に基づいて ESM データの偏りについて論じた. ここで, 性格と心理状態をそれぞれ以下のように測定した. これらを総じて以降では個人特性と呼ぶ.

性格 Big-Five パーソナリティ [21] で定義される以下の 5 つの指標を質問紙により計測した: 神経症傾向, 外向性, 開放性, 調和性, 誠実性.

心理状態 従来研究 [8] において心理的幸福感との関連が深いとされている以下の 4 つの指標を質問紙により測定した: うつ傾向 [15], 主観的ストレスレベル [6], 孤独感 [24], 人生満足度 [9].

また, 本研究では, ESM に対する回答の時間的な変化を以下のように算出した.

ESM 回答の時間的な変化 以下の 4 つの指標を実験期間中の ESM 回答から測定した. これらの指標は, ESM に対する回答頻度に基づいて図 1 のように抽出される.

- 継続性: 実験開始日から最後の ESM 回答をするまでの日数
- 平均頻度: 1 日あたりの ESM への平均質問回答数
- トレンド: 実験期間全体における 1 日あたりの ESM への質問回答数の増加/減少傾向
- 区間トレンド: 実験期間の一部における 1 日あたりの ESM への質問回答数の増加/減少傾向

我々は本研究課題を解決するため, 個人特性と継続性, 平均頻度, トレンド, そして区間トレンドとの関係についてそれぞれ分析した.

本分析において, 我々は StudentLife データセット [39] を利用した. StudentLife とは, 精神的健康状態 (うつ, 孤独感, など), 学業成績 (GPA, テストの点数, など), そして日常行動 (ジムに行く, 友人と電話する, など) の関係を調査するために, 米国ダートマス大学の学生 48 名を対象に, 性格や心理状態に関するアンケート調査や携帯電話センサログ (GPS, 会話時間, など) の取得, ESM データ (気分, 授業の様子, など) の収集を 10 週間にわたって実施した研究プロジェクトである. このデータセットには,

¹ 日本電信電話株式会社 NTT 人間情報研究所
NTT Human Informatics Laboratories, NTT Corporation,
Yokosuka, Kanagawa 239-0847, Japan

^{†1} 現在, 横浜市立大学データサイエンス学部
Presently with School of Data Science, Yokohama City University

a) tomu.tominaga.hm@hco.ntt.co.jp

b) shuhei.yamamoto.ea@hco.ntt.co.jp

c) takeshi.kurashima.uf@hco.ntt.co.jp

d) hirotoda@acm.org

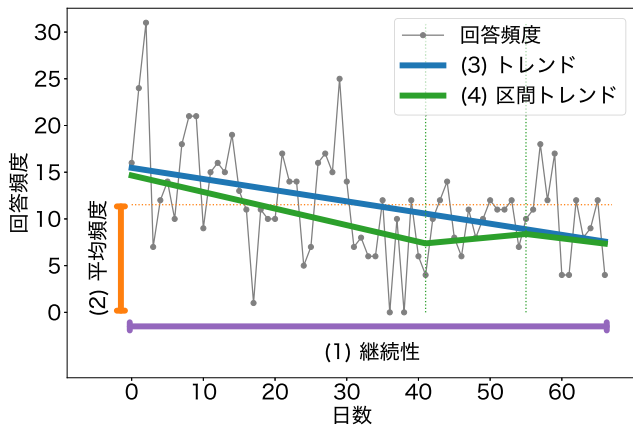


図 1 ESM 質問回答数の時間的変化の測定尺度（継続性、平均頻度、トレンド、区間トレンド）の抽出例

Fig. 1 A sample case of how to extract four measures of temporal response patterns from daily responses to ESM. The trend (third measure) is calculated from daily frequency. For the period-separated trend (fourth measure), three periods are given: week #1–#6 (6 weeks), week #7–#8 (2 weeks), and week #9–#10 (2 weeks). The fourth measure is calculated from average daily frequency per week.

大学で行われた中間試験や学校祭といった学校行事の情報も含まれる。本分析で我々は、本データセットに記録されている被験者の質問紙調査に対する回答結果をもとに被験者の性格や心理状態の測定尺度を算出した。また、本データセットには ESM に対する各被験者の回答ログが時系列に記録されているため、それらの情報を利用して ESM 回答数の時間的変化を算出した。

我々は、継続性、平均頻度、およびトレンドを個人特性から説明するための回帰分析を実施した。また、区間トレンドについては、中間試験や学校祭という週単位の学校行事の影響を考慮して実験期間を 3 区間に分割し、各時区間内におけるトレンドと、各トレンドが個人特性によってどの程度説明されるかを潜在成長曲線モデルを用いて分析した。

主な分析結果として、ストレスレベルの高い被験者の ESM 質問回答数は、実験期間全体を通じて多いものの、中間試験後に強く減少する傾向が確認された。また、内向的な（外向性の低い）被験者については、実験期間を通じて回答数の減少傾向は緩やかだが、学校祭後に強く回答数が減少することが分かった。我々は分析結果をふまえて、ESM 回答データが特定の個人特性や時間に偏りやすくなることを議論し、将来的に ESM を利用する研究が注意すべき点やこのような問題に対処するための実験設計について述べた。なお、本論文は著者らの先行研究 [33] をもとに投稿を行ったものである。

2. 関連研究

ESM は、人々の感情や思考、経験などの主観的な状態を時系列に記録するための有力な方法として認められている。ESM の特徴は、実験を実施する者が携帯端末を通じて被験者に定期的に質問を送り、被験者はその時その場でリアルタイムにその質問に答える点にある。そのため、ESM は従来の質問紙調査よりも想起バイアスが少なく、生態学的妥当性が高いデータを取得できる。ESM で収集されたデータは、行動パターンの特特定 [11], [25], [39], 人生のアウトカム予測 [8], [41], また人間関係の理解 [10], [13] など、多様な用途で利用されている。

しかしながら、ESM はその特性上様々な要因の影響を受けるため、特定の条件にあてはまるデータが偏って集められてしまうことがある [19]。このような影響を及ぼす要因は、被験者要因、文脈要因、設計要因の 3 つに分けられる。被験者要因とは、ESM の対象である被験者の個人的な性質を指しており、属性情報 [20] や心理特性 [27]、精神的健康状態 [14] などが含まれる。たとえば、男性 [20]、若年層 [20]、アルコール消費が多い [27]、負の感情や精神疾患をかかえている [19], [27]、といった項目にあてはまる被験者の回答数は少ない傾向にあることが報告されている。Vachon らのメタ分析 [36] でも、ESM の被験者に男性や精神疾患を持つ人が多く含まれるほど収集される回答数が少ない傾向にあることが示されている。

文脈要因とは、ESM の被験者の周囲の状況や環境に関する要因であり、時間や場所 [3]、携帯端末の設定 [37] などが該当する。ESM に対する回答データの時空間分布を調査した研究によれば、被験者は朝よりも夜中に、移動中よりも病院やレストランといった場所から回答する傾向にあることが示された [3]。また、スマートホンの状態（バッテリー残量、通知履歴、など）と ESM 回答率の関係を分析した研究 [37] は、（実験とは無関係の）通知を受け取る回数と ESM 回答率は正の相関関係にあるが、全体として回答数が少ない被験者に限ると、多くの通知を受け取ると反対に ESM 回答率が下がる傾向にあったと報告した。

設計要因とは ESM の実験設計に関わる要因であり、インセンティブ [20], [36] や質問のタイミングや頻度 [27], [38]、サンプリング方法 [36] などの因子があげられる。たとえば、ESM の質問を被験者の携帯端末のスクリーンがアンロックされたときに送ると、質問をランダムなタイミングもしくは一定の頻度で送る場合と比べて多くの回答が得られることが分かっている [38]。しかし、携帯端末のセンサーログに基づいて ESM 質問を送付するタイミングを制御することもまた ESM データに偏りをもたらし場合がある。携帯端末の加速度センサ、GPS、マイクセンサ、スクリーン情報、電話や E メールに関するログに基づいて質問の

タイミングを調整した研究によれば、全回答データのうち25%が午前10時から午後1時の間のものだったと述べられている [17]。これは、ランダムに ESM の質問を送付した場合に夜中の回答率が高いことを報告した研究 [3] とは異なる結果を示しており、実験設計の違いによって生まれた偏りであると考えられる。

上述のような ESM の回答データの偏りを調査する従来研究において、回答数の時間的な被験者内変動はほとんど注目されていない。また、回答数の時間変化と個人特性の関係性についても検証されていない。すなわち現状では、どの被験者のどのタイミングの回答データが少ないか（もしくは多いか）、という疑問に答えられないといえる。これが明らかになれば、我々は個人特性という観点に加えて時間という観点で ESM データの偏りが生じる条件を記述できるだけでなく、特定の被験者に対して特定のタイミングで回答を催促するための実験設計の指針を提示できる。そこで本研究は、個人特性が ESM 質問回答の時間変化とどのような関係にあるかを明らかにし、ESM データの個人特性や時間に対する偏りについて新たな知見を獲得する。

3. データセット

本研究課題に取り組むため、我々は StudentLife データセット [39] を活用した。StudentLife [39] は、個人特性を取得するためのアンケート調査、被験者の状況や経験を把握するための ESM 調査、そして日常行動を観測するためのスマートフォンセンサログの自動収集を、48名の米国ダートマス大学の学生に対して10週間にわたって実施したプロジェクトである。得られたデータは StudentLife データセットとして第三者向けに公開されている*1。本研究の分析において、我々はアンケート調査で取得された個人特性データと ESM 調査で収集された ESM 回答データを利用した。我々の知る限り、被験者の個人特性と ESM 回答履歴の対応関係が記録されており、かつ第三者向けに公開されているものは StudentLife データ以外にない。これらの詳細について以下に示す。

3.1 個人特性データ

ESM 回答数は、被験者のモチベーションが高ければ多く、低ければ少なくなるといった変化を示すと想定される。たとえば、誠実で勤勉な被験者はそうでない被験者よりも継続的に多くの回答を提出する傾向にある可能性がある。そこで我々は、被験者の性格を表現する心理尺度として学術的に広く認められている Big-Five パーソナリティ [21] を個人特性として位置付けた。具体的には、Big Five パーソナリティ [21] の要素である神経症傾向、外向性、開放性、調和性、誠実性を利用した。

また、被験者のモチベーションは実験期間中の出来事やイベントにも左右され、結果的に ESM 回答数を変化させる。特に、精神的に不調な被験者の回答数が少ないという従来研究の報告 [19], [27] をふまえると、ネガティブな出来事に敏感に反応したり、それによって精神的に不安定になったりする被験者はそのようなタイミングで回答を避ける可能性がある。反対に、心理的に充実した状態にある被験者は上記のような経験があっても回答を継続できると想定される。このような現象をとらえるため、被験者の心理的・精神的な充足度を測定する尺度として心理的幸福度を個人特性として採用した。具体的には、先行研究 [8] で心理的幸福度の構成要素であるとされているうつ傾向 (PHQ-9) [15]、主観的ストレスレベル (Perceived Stress Scale) [6]、孤独感 (UCLA Loneliness Scale) [24]、人生満足度 (Flourishing Scale) [9] を利用した。

これらのアンケート調査は、被験者個人に紐づく属性としての性質を測定する目的で行われている。すなわち、このアンケートに対する回答結果は、被験者の個人特性を表現するものとして解釈できる。

このアンケート調査は、実験開始時と終了時にそれぞれ実施されているが、我々は実験開始時の回答結果を用いた。我々は、本データセットに含まれる48名のうち、上記の個人特性のいずれかにおいてアンケートのすべての質問項目に回答していない被験者6名を除いた。分析対象となった42名の被験者の個人特性に関する基本的な統計量を表1の上段に示す。

3.2 ESM 回答データ

StudentLife プロジェクトで実施された ESM 調査では、気分やストレス、睡眠、授業など学生生活の日常における様々な側面に関する質問を1日あたり平均で8回程度の頻度で学生に送り、回答を収集している。具体的には、この ESM 調査では事前に26種類の検査項目（気分、ストレス、など）が用意されており（1検査項目あたり1~10問、計81問の質問を含む）、そのうち3~13種類の検査項目が1日のうち数回（疑似）ランダムなタイミングで被験者に配布した実験用スマートフォンにポップアップ表示される。回答時間に制約はなく、被験者は任意のタイミングで回答できる。先行研究 [39] によれば、被験者は気分やストレスといった質問文の短いもしくは質問数の少ない項目（文献 [39] 図3参照）に多く回答していたことが確認されている。明示的なデータや報告は確認されていないが、これらの質問は簡素に設計されているため、回答に要した時間はおよそ数分程度もしくはそれ未満であったと推察される。ESM に多く回答した被験者は、活動量計リストバンドやスマートタブレットを報酬として受け取った。詳細は先行研究 [39] の記述を参考にされたい。

この ESM 調査は、被験者の日常的な生活や環境を把握

*1 <https://studentlife.cs.dartmouth.edu/dataset.html>

表 1 分析対象データの基本統計量 (N = 42)

Table 1 Descriptive statistics of data for our analysis.

	最小値	平均	最大値
人格			
神経症傾向	14	23.78±5.63	36
外向性	13	23.40±6.33	36
開放性	25	35.88±4.85	46
調和性	12	32.83±5.62	41
誠実性	17	30.85±6.45	43
心理状態			
うつ傾向	1	2.25±0.98	5
主観的ストレスレベル	9	20.60±3.99	28
人生満足度	25	41.20±10.90	64
孤独感	16	42.40±7.87	54
ESM 回答データ			
回答数	30	436.64±301.85	1585
回答日数	4	58.10±17.26	67

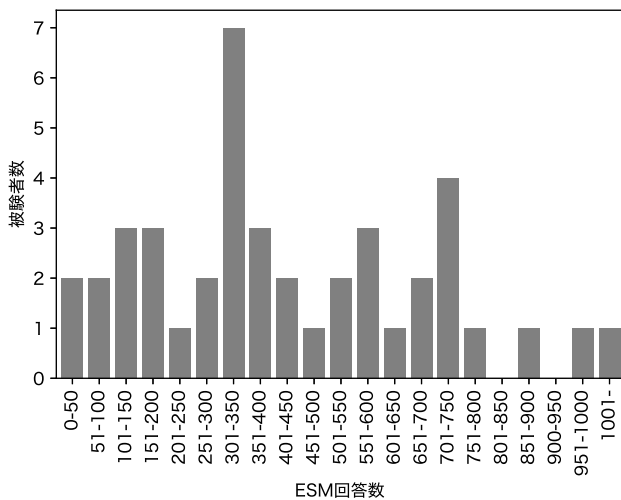


図 2 ESM 回答数に対する被験者数の分布

Fig. 2 A distribution of subjects to ESM responses.

する目的で、実験期間を通じて継続的に（毎日）実施されたものである。そのため、調査の様式としては個人特性のアンケート調査と同様に自己報告式ではあるが、測定する対象や方法は大きく異なる点に注意されたい。たとえば、個人特性のアンケート調査において測定される主観的ストレスレベルとは、被験者がそもそもどの程度ストレスをかかえやすいかという個人的性質を事前に測定した結果であり、ESM 調査において測定されるストレスとは、被験者が日々の中で体験したできごとに対してどの程度ストレスを持ったかという日常的経験を即時的に観測した結果である。

本分析では、ESM 回答傾向に対して個人特性がどのように影響しているかを知ることが目的であるため、個人特性とは無関係にそもそも実験への動機付けが極端に低い被験者がいた場合、それらを分析に含めない方がよい。そこで、表 1 に示される 42 名の被験者の回答数の分布を確認したところ、図 2 のようになり、対象の中で極端に回答数

の少ない被験者の存在は確認されなかった。ゆえに、対象とした 42 名はこの ESM 調査において一定の動機付けがなされていたと考えられる。これは前述のように、活動量計リストバンドやスマートタブレットといった高額な報酬が提示されていたことによる効果と思われる。以上から、本分析では対象被験者が 10 週間で回答した 18339 件の ESM 回答データをすべて利用した。42 名の被験者の ESM 回答に関する統計量を表 1 の下段に示す。

4. 分析

我々は ESM 回答の継続性、平均頻度、およびトレンドと個人特性の関係を調べるために重回帰分析を実施した。また、区間トレンドと個人特性の関係については潜在成長曲線 (Latent Growth Curve, 以下 LGC) モデルを活用して分析した。以下にそれぞれの詳細について述べる。

4.1 継続性、平均頻度、およびトレンド

継続性、平均頻度、もしくはトレンドを目的変数、個人特性の各尺度を説明変数とした重回帰分析を行った。説明変数は平均 0、分散 1 となるように事前に標準化を実施した。この分析によって得られた各説明変数の目的変数に対する偏回帰係数を、個人特性が継続性、平均頻度、およびトレンドに及ぼす影響と解釈した。本論文において我々は、統計的な有意性が認められた偏回帰係数に基づいて個人特性の影響力を議論した。

4.2 区間トレンド

LGC モデルは対象データの経時変化をサンプルの全体レベルおよび個人レベルでとらえることができる。具体的には、LGC モデルによって、誰が、いつ、どのような時間変化を示したかを把握できる。そこで我々は、ESM 回答数の時間変化を LGC モデルに適用し、区間トレンドと個人特性との関係性を調査した。

4.2.1 観測変数と潜在変数

LGC モデルとは、主に縦断データを対象に、共分散構造分析のフレームワークを利用して観測変数の経時的変化 (成長) を明らかにする統計的解析手法の一種である。本分析では、各週における 1 日あたりの平均 ESM 回答数を観測変数とする。これは、後述のように本分析のサンプルである学生が所属する大学における校内行事 (中間試験と学校祭) が週単位で実施されている影響を考慮し、ESM 回答数を週レベルに集約したためである。

LGC モデルにおいて、観測変数の経時的変化は切片 (初期値) と n 次の成長率で記述される。本分析では、より簡潔なモデルで現象を記述することで解釈性を担保することを目的として、成長率を 1 次 (線形変化) に設定した。各サンプルの切片と成長率は LGC モデルで潜在変数として導入され、推定される。

4.2.2 観測期間の分割

LGC モデルは、観測変数のより詳細な経時的変化をとらえるために、観測期間を分割し各時区間における成長率を推定できる。各時区間における成長率を以降では区間トレンドと呼ぶ。観測期間の区切り方は、対象データの時間変化を目視で確認する、もしくは事前知識をふまえるといった明示的な方法で行われることが多い。たとえば、特定の時点で明らかに対象データが異なる傾向を見せる場合や影響力の大きいイベントが起きた場合には、そのタイミングで観測期間を区切る。

本分析において我々は、LGC モデルを適用する際に観測期間を第 6 週と第 8 週を境界として 3 つに分割した。各時区間を第 1 区間（第 1 週～第 6 週）、第 2 区間（第 6 週～第 8 週）、第 3 区間（第 8 週～第 10 週）と呼ぶ*2。このように観測期間を分割した理由は 2 つある。

第 1 に、ESM の質問への回答頻度は中間試験と学校祭に影響されうるからである。StudentLife データセットを用いた過去の分析 [11], [40] は、日常的な行動パターンが上記のイベントの前後で変化したことを指摘しているため、本研究においても同様の傾向が確認される可能性が高いと我々は考えた。中間試験は第 3 週から第 5 週にかけて、学校祭は第 7 週に実施されたことを考慮して、我々は観測期間を第 6 週と第 8 週で分割した。

第 2 に、第 6 週と第 8 週で分割したときのモデルが高い適合度（4.2.5 項参照）を示したからである。第 1 の理由をふまえると、実験期間を分割する方法は、第 6 週もしくは第 8 週を境界として 2 分割する場合、もしくは第 6 週と第 8 週を境界として 3 分割する場合の 3 つが考えられる。我々は、これらの分割によって得られる 3 つの LGC モデルの適合度を調べ、第 6 週と第 8 週による 3 分割で得られるモデルが最も高い適合度を示したことを確認した。

4.2.3 モデルの構造

上記の設定に基づいて構成された LGC モデルにおける観測変数と潜在変数の関係を図 3 (A) に示す。ここで、観測変数である第 t 週目（ $1 \leq t \leq 10$ ）の 1 日あたりの平均 ESM 回答数を x_t 、潜在変数である切片を h_0 、第 i 区間（ $1 \leq i \leq 3$ ）における区間トレンドを h_i として、上記のモデル構造は以下のように記述される。

$$x_t = \epsilon_t + \begin{cases} h_0 + (t-1)h_1 & (1 \leq t \leq 6) \\ h_0 + 5h_1 + (t-6)h_2 & (6 < t \leq 8) \\ h_0 + 5h_1 + 2h_2 + (t-8)h_3 & (8 < t \leq 10) \end{cases} \quad (1)$$

ここで ϵ_t は誤差項を示す。図 3 (A) のモデル（式 (1)）を以降では非条件 LGC モデルと呼ぶ。

*2 第 6 週と第 8 週が 2 つの時区間に含まれている理由は、LGC モデルにおいて第 2 区間以降の各時区間における最初の週はその時区間における切片であり各区間トレンドから 0 の影響力を受ける、と見なすためである。

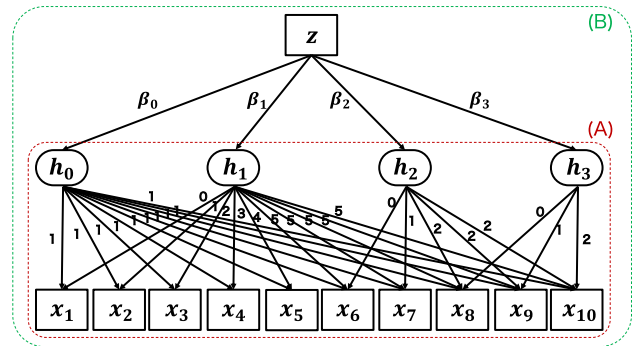


図 3 本分析における潜在成長曲線モデルの構造の概略 (h_i : 第 i 区間 ($0 \leq i \leq 3$) における区間トレンドを表現する潜在変数 (ただし h_0 は切片を表す), x_t : 第 t 週 ($1 \leq t \leq 10$) の 1 日あたりの平均 ESM 回答数を表現する観測変数, z : ある個人特性を表現する観測変数, β_i : 個人特性 z の潜在変数 h_i に対する回帰係数). (A) 非条件 LGC モデル: サンプル全体の時間変化 (各サンプルの h_i の平均) を推定するためのモデル, (B) 個人特性条件 LGC モデル: 区間トレンドに対する個人特性の影響 (各サンプルの h_i と z の関係) を分析するためのモデル

Fig. 3 Conceptual structure of latent growth curve models in our analysis (h_i : latent variables of intercept and trend in i -th period ($0 \leq i \leq 3$), x_t : daily average ESM responses in t -th week ($1 \leq t \leq 10$), z : some personal characteristic, β_i : a partial regression coefficient of z for h_i). The latent variables are in rounded rectangles, and the observed variables are in rectangles. (A) Unconditional model, estimating overall trend of temporal ESM response patterns for all subjects. (B) Conditional model, introducing a factor of personal characteristics to the unconditional model to infer effects of the personal characteristic on temporal ESM response patterns.

非条件 LGC モデルは、共分散構造分析（因子分析）と同様に潜在変数を推定するため、各サンプルの各潜在変数の因子得点を算出する。したがって、この因子得点の平均値が対象サンプル全体の時間変化の傾向に相当し、その分散が対象サンプル間の時間変化のばらつきに相当する。

4.2.4 個人特性の導入

LGC モデルは、サンプルの持つ性質を新たな観測変数（説明変数）として導入し潜在変数との関係を回帰分析することで、潜在変数の分散がサンプルの性質からどの程度説明されるか、すなわち、どのような被験者がどのような時間変化を示すかを明らかにする能力を持つ。本分析では、個人特性の尺度をそれぞれ説明変数として導入し、各潜在変数に対する偏回帰係数を算出した。この分析で利用されるモデルにおける潜在変数、観測変数、および説明変数の関係を図 3 (B) に示す。ここで、ある説明変数を z 、各潜在変数 h_i に対する偏回帰係数を β_i とすると、説明変数と潜在変数の関係は以下のモデル式で表現される。

$$h_i = \beta_i z + \epsilon_i \quad (2)$$

ここで ϵ_i は誤差項を示す。図 3 (B) で記述されるこのモ

デル構造 (式 (1) および式 (2)) を以降では個人特性条件 LGC モデルと呼ぶ。

4.2.5 モデルの適合度

それぞれの LGC モデルは上記の設定において対象データに適合するように潜在変数を推定する。モデルの適合度は近似誤差二乗平均平方根 (Root Mean Square Error of Approximation, 以下 RMSEA) [28], 比較適合度指標 (Comparative Fit Index, 以下 CFI) [1], タッカー・Lewis 指数 (Tucker-Lewis Index, 以下 TLI) [34] で評価される。RMSEA は低ければ低いほど, CFI と TLI は高ければ高いほどモデルが対象データによく適合していることを示す。

5. 結果

5.1 個人特性と継続性, 平均頻度, およびトレンドの関係

図 4 に各個人特性の継続性, 平均頻度, およびトレンドに対する偏回帰係数を示す。ここで Intercept とは各回帰モデルの切片を指す。各個人特性の値が 0 であるときの値に相当するため, 全体平均と考えることができる。たとえば, 被験者全体の平均として, ESM 回答数は 1 日ごとに 0.12 件ずつ減少したということが分かる。以降では, 個人特性の偏回帰係数に焦点を当てて結果を述べる。

主観的ストレスレベルは平均頻度に対して有意な正の偏回帰係数を示した ($\beta = 119.23 \pm 57.06^*$)。すなわち, 主観的ストレスレベルの高い被験者ほど実験期間全体を通じて多くの ESM 回答を行ったといえる。また, 外向性, 誠実性, そして主観的ストレスレベルのトレンドに対する偏回帰係数は有意な負の値を示した ($\beta = -0.035 \pm 0.016^*$, $-0.034 \pm 0.015^*$, $-0.033 \pm 0.015^*$)。この結果は, 被験者の外向性, 誠実性, もしくは主観的ストレスレベルが高ければ高いほど, そのような被験者の ESM 回答数はより強く減少する傾向にあることを意味している。

継続性については, 個人特性との間に有意な関係が確認されなかった。これはつまり, 被験者の個人特性は ESM に回答する期間の長さを説明するうえで有効でないことを示している。

5.2 個人特性と区間トレンドの関係

図 5 に LGC モデルを用いた分析結果を示す。各図において横軸は LGC モデルの種類を示す。たとえば, 「神経症傾向」とは, 非条件 LGC モデルに神経症傾向を個人特性の観測変数として導入したモデルを指す。LGC モデルが算出した潜在変数のうち, 最左列には切片, 左中列には第 1 区間トレンド, 右中列には第 2 区間トレンド, 最右列には第 3 区間トレンドに関する結果が示されている。上段と中段には各モデルにおける各潜在変数の平均値と分散の値が示されている。また, 下段には個人特性条件 LGC モデルにおける各個人特性の潜在変数に対する偏回帰係数が示されている。なお, 個人特性条件 LGC モデルは, 非条件

LGC モデルに個人特性を新たに説明変数として導入し潜在変数のばらつきを説明するような偏回帰係数を算出するため, 非条件 LGC モデルと比較した際に, 対象である潜在変数の分散の減少が期待される。一方で, 平均値には大きな影響がないため, 以降において各潜在変数の平均値について言及する際には非条件 LGC モデルの結果を参照する。

非条件 LGC モデルにより算出された潜在変数の平均値の結果に見られるように, すべての区間トレンドが負の値を示した ($h_1 = -0.96$, $h_2 = -0.84$, $h_3 = -0.79$)。これは, トレンドが全体として負の値を示したこと (5.1 節) と一貫する結果である。

主観的ストレスレベルは切片に対して正の, 第 2 区間トレンドに対して負の偏回帰係数を持つことが分かった ($\beta_0 = 1.88^*$, $\beta_2 = -0.68^{**}$)。非条件 LGC モデルと比べて, 主観的ストレスレベルの個人特性条件 LGC モデルにおける潜在変数の分散がそれぞれ下がっていることも確認された (切片/非条件: $V = 25.79$, 個人特性条件: $V = 22.34$; 第 2 区間トレンド/非条件: $V = 2.03$, 個人特性条件: $V = 1.49$)。前者の結果は, 主観的ストレスレベルの高い被験者は初期値としての ESM 回答数が多いことを意味しており, 主観的ストレスレベルの高さと実験期間全体の ESM 回答数が正に相関する結果 (5.1 節) と一貫する。後者の結果は, 被験者の主観的ストレスレベルが高ければ高いほど, 第 2 区間すなわち中間試験終了後の ESM 回答数の減少傾向が強いことを意味する。

また, 外向性の第 3 区間トレンドに対する偏回帰係数が有意な正の値を示した ($\beta_3 = 0.52^*$)。非条件 LGC モデルと比較して外向性を導入した個人特性条件 LGC モデルの方が低い分散の値を示したことも確認された (非条件: $V = 2.26$, 個人特性条件: $V = 1.95$)。これはすなわち, 外向性が低い (内向性が高い) 被験者は, 第 3 区間すなわち学校祭の後に ESM 回答数が強く減少する傾向にあることを意味する。

さらに, 調和性が第 1 区間トレンドに対して有意に正の偏回帰係数を持つことが分かった ($\beta_1 = 0.31^*$)。これは, 実験の初期段階において, 調和性の高い被験者の ESM 回答数の減少傾向が弱いことを示している。

図 6 にモデル適合度の結果を示す。一般に, CFI と TLI は 0.90 以上, RMSEA は 0.10 未満であればよいとされているが, 本分析におけるすべての LGC モデルが RMSEA で 0.10 以上の値を示しており, モデルが対象データに十分適合しているとはいきれない。これは, 特にサンプル数の少なさの影響を受けていると思われる。この点については 6.4 節で議論する。

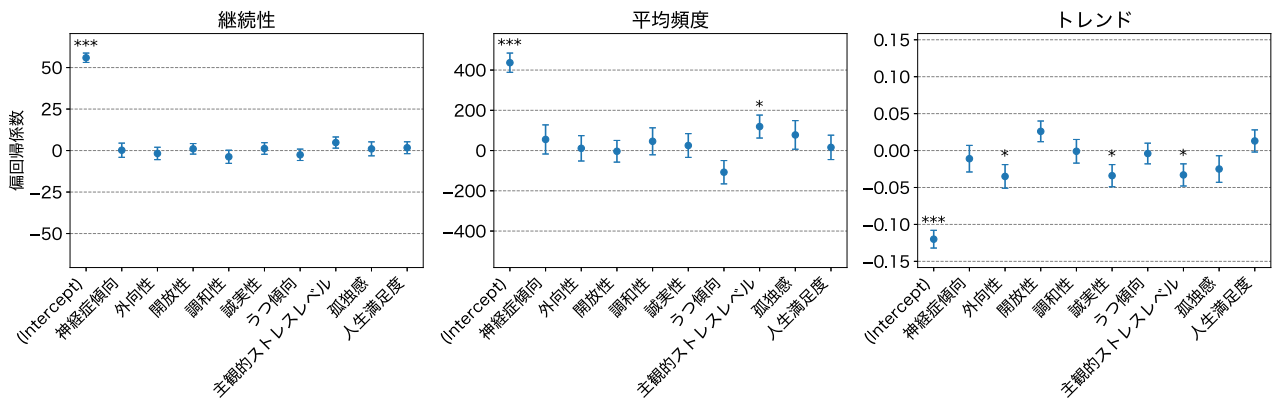


図 4 継続性，平均頻度，トレンドを個人特性から説明する回帰分析によって得られた偏回帰係数（エラーバーは標準誤差； p 値：***... $p < 0.001$, **... $p < 0.01$, *... $p < 0.05$ ）

Fig. 4 Regression models explaining continuity, frequency, and trend of ESM responses from personal characteristics (error bars represent standard errors).

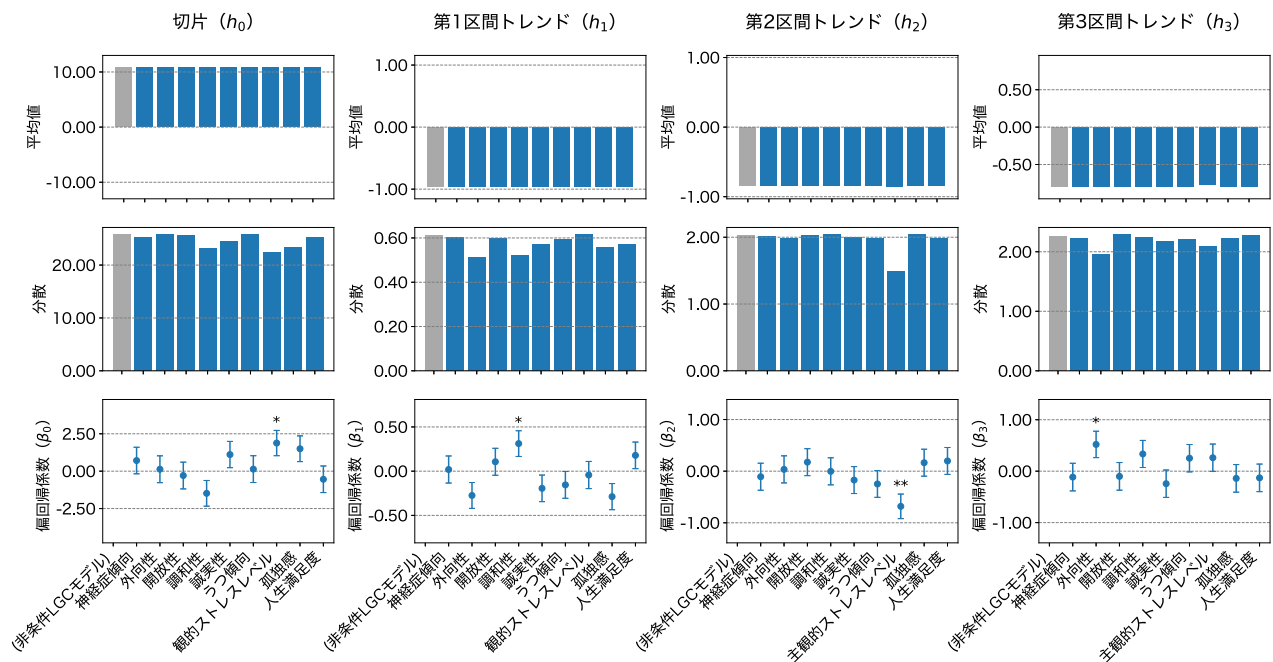


図 5 LGC モデルによって算出された ESM 回答数の時間変化に関する指標（切片および区間トレンドに関する潜在変数）と個人特性の関係（下段図におけるエラーバーは標準誤差； p 値：***... $p < 0.001$, **... $p < 0.01$, *... $p < 0.05$ ）

Fig. 5 Latent variables of temporal ESM response patterns and their relationships with personal characteristics. These values are estimated by unconditional and conditional LGC models. Error bars represent standard errors (p -value : ***... $p < 0.001$, **... $p < 0.01$, *... $p < 0.05$).

6. 考察

6.1 影響力のある個人特性

我々の結果で示唆されたように，主観的ストレスレベル，外向性，誠実性，調和性は実験期間全体の ESM 回答数やその時間変化と有意な関係を持つことが分かった。以下では，それらの関係が確認された理由について考察する。

6.1.1 主観的ストレスレベル

本研究の分析結果で見られたように，主観的ストレスレ

ベルを高いと報告した被験者は 1 日あたりの平均 ESM 回答数が多い (5.1 節)。先行研究 [18] によれば，ストレスをかかえる多くの方は，人に対してではなくコンピュータ上での自己開示を好む。ESM を利用した研究において，被験者は携帯端末で自身の気分や経験に関する ESM の質問に回答するという行為をコンピュータ上の自己開示ととらえている可能性がある。そのため，主観的ストレスレベルの高さが ESM 回答数の多さに繋がっていると考えられる。さらに，主観的ストレスレベルの高い被験者は実験期

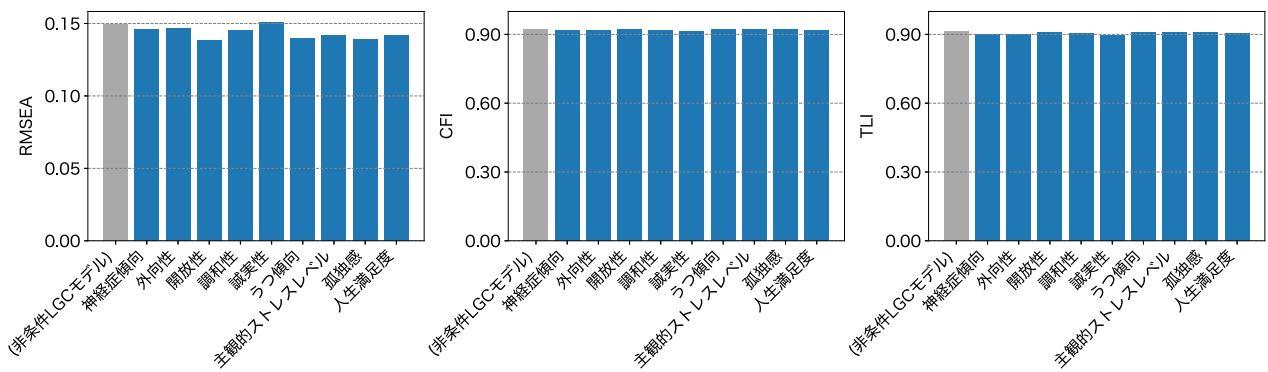


図 6 各 LGC モデルの適合度
Fig. 6 Goodness of fit of LGC models in the analysis.

間全体を通じて ESM 回答数が大きく減少する傾向にあり (5.1 節), 特に中間試験の後にその傾向が強いこと (5.2 節) が確認された. 先行研究 [11] で述べられているように, 中間試験に対して学生は大きなプレッシャーを感じる. 実際に, 被験者を主観的ストレスレベルの平均値で 2 群に分け, 第 1 区間で被験者が ESM で報告したストレスレベル (1: 低い~5: 高い) の変化を確認したところ, 図 7 に示されるように主観的ストレスレベルの高い被験者群は中間試験の実施期間中 (第 3~6 週) にかけてストレスをかかえていたことが分かった. そのような精神的な負荷はコンピュータ上での自己開示を促す可能性がある [18]. これらの点を考慮すると, 主観的ストレスレベルの高い被験者が中間試験の終了によって一時的な精神的負荷から解放されたことにより, コンピュータに対する自己開示すなわち携帯端末上での ESM 回答の頻度が下がったのではないかと考えられる.

6.1.2 外向性と内向性

我々は, 外向性が実験期間全体の ESM 回答数の減少傾向を強める (5.1 節) 一方で, 第 3 区間における減少傾向を弱めること (5.2 節) をそれぞれ確認した. これらは直感的には相反する結果のように見える. まず, これらの結果を解釈する.

図 5 に示されているように, 外向性を導入した個人特性条件 LGC モデルの各区間トレンドに対する偏回帰係数の値は $\beta_1 = -0.27, \beta_2 = 0.03, \beta_3 = 0.52$ である. これらの値から外向性が実験期間全体にわたる週レベルのトレンドに及ぼす効果の総和はおよそ $-0.27 (= -0.27 \cdot 5 + 0.03 \cdot 2 + 0.52 \cdot 2)$ となる. 図 4 に示されているように, 外向性のトレンドに対する偏回帰係数の値は -0.035 である. これらはいずれも一貫して負の値を示している. すなわち, 外向性は第 3 区間トレンドに対しては正の影響を及ぼす (第 3 区間における回答数の減少傾向を弱める) が, 各区間トレンドの値の大きさとその期間の長さを考慮すると実験期間全体としてはトレンドに対して負の影響を及ぼす (実験期間全体の回答数の減少傾向を強める) と解釈できる.

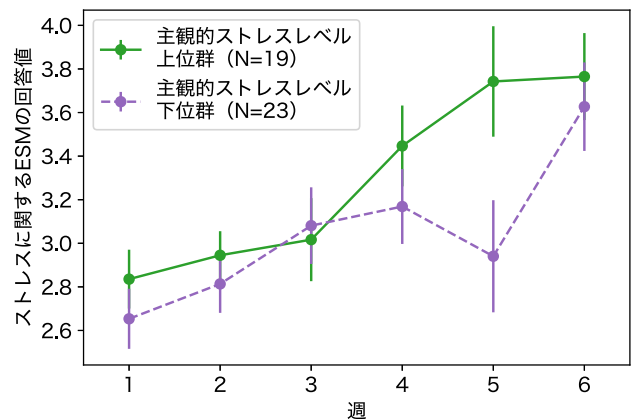


図 7 主観的ストレスレベル上位群 (N=19) と下位群 (N=23) の第 1 区間におけるストレスに関する ESM 質問への回答値 (1: 低い~5: 高い) の推移
Fig. 7 Weekly stress scores as assessed by ESM questions (1: low-5: high) during the first period for subjects with high (N=19) and low (N=23) perceived stress scale.

次に, 被験者の外向性が低ければ低いほど, すなわち内向性が高ければ高いほど第 3 区間トレンドの減少傾向が強くなる点について考察する. 内向的な性質を持つ人々は, 自身の内面的な側面について内省することを好む [32]. ESM 質問に回答する際には自身の気分や経験について考えるため, ESM 回答は被験者にとって内省に相当する行為としてとらえられている可能性がある. 先行研究 [31] によれば, このような内省には心理的な資源やエネルギーが必要となる. しかし, 心理的資源は社会的な相互作用によって消耗されてしまうことが指摘されている [12]. これらの点をふまえると, 内向的な被験者が学校祭というイベントにおける社会的な交流によって心理的資源を消耗し, その結果そのような被験者の自身の行動や経験に対する内省すなわち ESM 質問への回答の頻度が減少した可能性がある.

ここで, 内向的な被験者がどの程度社会的な交流を持っていたかを理解するために, 我々は StudentLife データセットに記録されている被験者の会話時間を分析した. この会話時間のデータは, 被験者の携帯端末のマイクセンサから

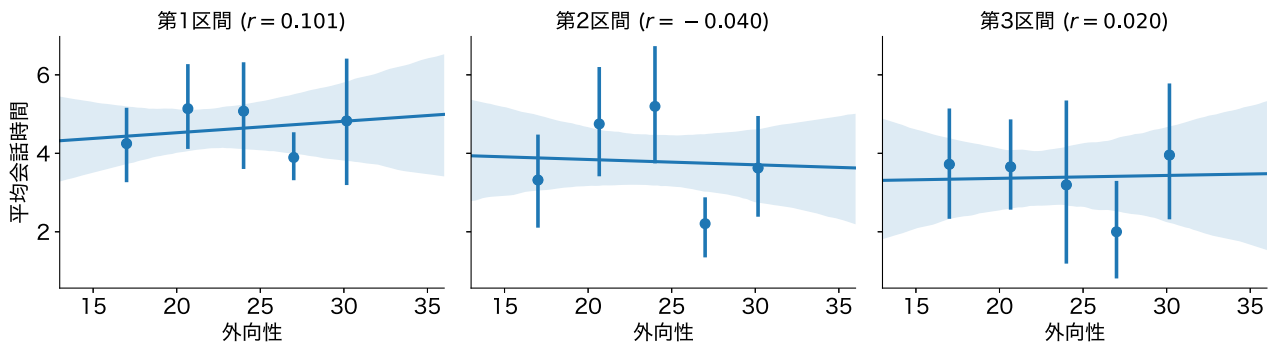


図 8 各時区間における 1 日あたりの平均会話時間と外向性の関係
 Fig. 8 Extraversion and daily average conversation duration (hours) in each period.

得られた音声データから会話か否かを自動で分類する技術 [16], [22] を用いて推定されたものである。我々はこのデータを社会的な活動を表す指標として採用した。

図 8 に、被験者の外向性と各時区間における 1 日あたりの平均会話時間の関係を示す。外向性の高い被験者が社会的な交流を好むことをふまえると、外向性と会話時間は正の相関関係を示すことが期待される。結果として、第 1 区間と第 3 区間では緩やかな正の相関が見られたが、第 2 区間では期待に反して緩やかな負の相関が確認された。これは、外向性の低い (内向的な) 被験者が期待よりも多く会話したことを意味する。特に、外向性の高さに基づいて各群に含まれる人数が等しくなるように被験者を 5 段階に (第 1 群の外向性が最も低く、第 5 群の外向性が最も高くなるように) 分けた場合、図 8 中央に示されるように、第 2 区間において第 2 群と第 3 群の被験者の会話時間が期待より高い値を示していることが分かる。したがって、学校祭が開催された期間で内向的な被験者は多くの社会的交流を行っていたことが示唆された。ただし、ここで示された相関関係に統計的有意性は見られなかったため、弱い相関関係として確認されたことに注意されたい。

6.1.3 誠実性

誠実性の高さは、実験期間全体を通じた ESM 回答数の減少傾向の強さと相関したが、平均頻度や継続性とは有意な関係になかった (5.1 節)。一般に、誠実な人は自律的または忠実であり、目標達成志向が高いことが知られている [7]。そのため、誠実性の高い被験者はより多くの ESM 回答をより長期にわたって実施することが期待されるが、本研究の結果はこれを支持するものではなかった。

この点について理解を深めるため、我々は誠実性の高さの上位 5 名と下位 5 名を対象に、1 日あたりの平均 ESM 回答数が週単位でどのように変化するかを比較した。その結果、図 9 に見られるように、上位 5 名の第 1 週目における ESM 回答数は下位 5 名に比べて多いことが確認された。これは、誠実性が高い被験者は実験の初期段階では多くの ESM 回答を行うものの、その後の期間はそうではないことを意味している。

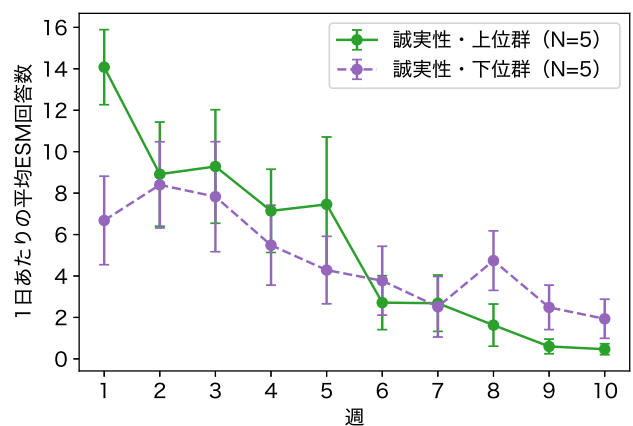


図 9 誠実性上位 5 名と下位 5 名における 1 日あたりの平均 ESM 回答数の週単位での遷移
 Fig. 9 Daily average ESM responses by week of top-5 and bottom-5 conscientiousness subjects.

また、中間試験を含む第 5 週から第 6 週にかけて上位 5 名の被験者の ESM 回答数が強く減少している。同一のデータセットを利用した従来研究は、誠実性の高い学生は良い学業成績を残したことを報告している [40]。そのため、誠実性の高い被験者は中間試験を重視し、ESM 質問に対する回答に時間を使わなかった可能性がある。これが、上記の期間における ESM 回答数の減少傾向につながったと考えられる。

つまり、誠実性の高い被験者は実験の初期段階においてのみ多くの ESM 回答を行ったが、中間試験で良い成績を残すことに注力するために中間試験期間中には ESM 回答を避けたことで ESM 回答数が減少する傾向にあり、その後は誠実性の低い被験者と同程度の頻度で ESM 回答を行ったといえる。そのため、誠実性は実験期間全体を通じた減少傾向を強める結果となったが、平均頻度や継続性に対しては有意な影響力を持つことはなかったと解釈できる。

6.1.4 調和性

調和性が高ければ高いほど、その被験者の第 1 区間における ESM 回答数の減少傾向は弱まること (5.2 節) が本分析で明らかになった。この理由を探るため、我々は被験者

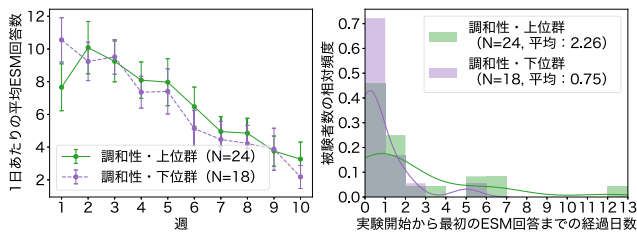


図 10 調和性が平均値より高い被験者群 (N=24) と平均値以下の被験者群 (N=18) の比較：(左) 1日あたりの平均 ESM 回答数の週単位の遷移を比較した。その結果、図 10 左に見られるように、調和性の高い被験者群は第 1 週から第 2 週にかけて ESM 回答数が増加することが分かった。これに対し我々は、調和性の高い被験者は最初の ESM 回答が調和性の低い被験者と比べて遅れているのではないかと考えた。

Fig. 10 Comparison of subjects with high agreeableness (N=24; those agreeableness is higher than the average) and subjects with low agreeableness (N=18; those agreeableness is lower than the average): (Left) daily average ESM responses by week, (Right) relative cumulative frequency of subjects to the number of days from the beginning of the study to the initial ESM response.

を調和性の平均値で 2 群に分け、1 日あたりの平均 ESM 回答数の週単位の遷移を比較した。その結果、図 10 左に見られるように、調和性の高い被験者群は第 1 週から第 2 週にかけて ESM 回答数が増加することが分かった。これに対し我々は、調和性の高い被験者は最初の ESM 回答が調和性の低い被験者と比べて遅れているのではないかと考えた。

このことを確認するため、我々は実験が開始された日から各被験者が最初に ESM 質問に回答するまでに要した時間を調和性の高い群と低い群で比較した。図 10 右に、各群の実験開始時から最初の ESM 回答までの日数に対する被験者数を示す。

結果として、最初に ESM 回答を行うまでに、下位群は平均で 0.76 日要するのに対し、上位群は 2.26 日要していることが分かった。また、相対頻度を確認すると、下位群の被験者群の約 72% が初日に回答している一方で、上位群は約 46% という結果が示された。さらに、上位群と下位群で最初の ESM 回答までの日数の平均値の差を t 検定で確認したところ、統計的に有意な差分が見られた ($t = 2.245$, $p < 0.05$)。したがって、調和性の高い被験者は最初の ESM 回答が調和性の低い被験者と比べて遅れているといえる。しかし、我々は調和性の高い被験者がなぜ最初の ESM 回答に遅れてしまうかについては分かっていない。この点の検証については今後の課題とする。

6.2 個人特性と時間による ESM データの偏り

これまで示された結果と考察をふまえ、ESM 回答データが個人特性と時間によってどのように偏るのかを整理する。この議論において我々は、ESM データの偏りを全体バイアスと局所バイアスの 2 種類に分け、それぞれ以下のよ

うに定義する。全体バイアスとは、実験期間全体を通じて収集された ESM 回答が特定の個人特性を持つ被験者に偏ることを指す。局所バイアスとは、時間の経過にもなつて、もしくは特定の時期やタイミングによって、収集された ESM 回答が特定の個人特性を持つ被験者に偏ることを指す。本研究で得られた結果から示唆される ESM データの偏りをこれらの観点で整理したものを表 2 にまとめる。

本研究で対象とした個人特性の中で全体バイアスの要因となったものは主観的ストレスレベルのみであり、その他の個人特性は全体バイアスの要因としては認められなかった。我々の結果は、ESM データには全体として主観的ストレスレベルの高い被験者からの回答に偏ってしまう可能性を示唆している。

局所バイアスのうち、時間の経過にもなう偏りをもたらす要因として主観的ストレスレベル、外向性、誠実性が確認された。これらの個人特性が高ければ高いほどそのような被験者の回答数は時間の経過にもなう大きく減少していくため、実験の前半ではこれらの被験者の回答数は多いものの後半では少なくなるということを意味する。

また、特定の時期やタイミングによって引き起こされる局所バイアスも確認された。まず、ストレス性の高いイベントを経験した後に主観的ストレスレベルの高い被験者の回答は見込まれないことが分かった。次に、社会的な交流イベントの後には、内向的な被験者が ESM 回答を避ける傾向にあることを確認した。さらに、実験の初期段階では調和的な被験者の回答数が少なく、調和性の低い被験者の回答が多く集まる可能性が示唆された。

全体バイアスの結果は、被験者の個人的性質と ESM への回答との関係を調査したという点において、先行研究 [19], [23], [27], [36] と同様の分析から得られたものとして位置づけることができる。ただし、主観的ストレスレベルに焦点を当てた研究はほとんどないため、全体バイアスの結果は知見として新しいといえる。さらに重要なことは、局所バイアスの結果が個人特性による ESM データの不均一性が時間の経過やタイミングによってどのように異なるかを示した点である。これは従来研究にない観点の調査結果であり、本研究の貢献である。

6.3 ESM 研究における推奨事項

本研究の主たる貢献は、ESM 回答傾向が個人特性や時間に依存することを実証的に示した点にある。本節では、本分析によって得られた発見事項に基づき、ESM 研究が配慮すべき項目について論じる。

まず、将来的な ESM 研究は、被験者の個人特性と ESM 回答数の時間的変化の関係を記述し、知見として研究領域に蓄積すべきである。特に本分析からは主観的ストレスレベル、外向性、誠実性、調和性が影響力のある個人特性として特定されたため、これらを実験において考慮することを

表 2 個人特性と時間による ESM データの偏り
 Table 2 Bias in ESM data due to personal characteristics and time.

種類	個人特性	時間の経過/時期・タイミング	ESM 回答数の偏り方
全体バイアス	主観的ストレスレベル	-	主観的ストレスレベルが高いと多い
局所バイアス	主観的ストレスレベル	時間の経過	主観的ストレスレベルが高いと時間経過にともなう減少傾向が強い
	外向性 (内向性)	時間の経過	外向性 (内向性) が高いと時間経過にともなう減少傾向が強い (弱い)
	誠実性	時間の経過	誠実性が高いと時間経過にともなう減少傾向が強い
局所バイアス	主観的ストレスレベル	ストレス性の高いイベント	主観的ストレスレベルが高いとストレス性の高いイベント (定期試験など) の後の減少傾向が強い
	外向性 (内向性)	社会的な交流イベント	外向性 (内向性) が高いと社会的な交流イベント (学校祭など) の後の減少傾向が弱い (強い)
	調和性	実験の初期段階	調和性が高いと実験の初期段階における回答数の増加傾向が強い (最初に回答するまでが遅い)

推奨する。本研究では、実験期間中に被験者の行動パターンを変化させるようなイベントを被験者全員が共通して経験した点を考慮して、同一のデータセットを用いた先行研究 [11] を参考に LGC モデルを採用しているが、データの偏りを記述する方法はこれに限らない。たとえば、個人特性に応じて被験者を複数群に分けそれぞれの回答数の時間的推移について、各時点の単純な統計量を示すといった基本的な手法から、回帰不連続デザインや差分の差分法などの因果推論に用いる統計的な手法を用いることで、個人特性の違いがどのように回答の時間推移に影響を及ぼし、それがどのようなタイミングで起きたかを知ることができる。このような手続きによって記述された知見が蓄積されれば、各報告は事例研究であったとしても、将来的にメタ分析によって汎用的な知識を抽出できる可能性がある。

このような手続きによって得られた知識は、ESM データセットを利用した研究の実験手法を設計するうえで有用である。たとえば、携帯端末のセンサデータから個々人の将来の気分を予測するモデルを構築する ESM 研究において、主観的ストレスレベルによってもたらされる時間の経過に関する局所バイアスがデータセット内に存在する (つまり、その ESM の実施期間の前半にはストレスレベルの高い被験者の報告が多く、後半にはストレスレベルの低い被験者の報告が多い) 場合を考える。仮に、このことを事前に知らない場合、前半のデータをモデルの学習に利用し、後半のデータを評価・検証に利用してしまうと、根本的に (被験者の) 性質の異なるデータをそれぞれで用いてしまうため、正当にモデルの性能を確認することができない。反対に、このようなデータの偏りが事前に把握されていれば、研究実施者はその ESM データセットに含まれる特定の期間や被験者を除外する、もしくは別の ESM データセットを利用する、といった対応をとることが可能とな

り、データの偏りによる影響を受けない検証結果を得ることができる。

また、ESM データをなるべく偏りなくより多くの回答を収集するための実験設計を考案する際には、本研究の知見が有用である。たとえば、全体バイアスを避けるために、事前に主観的ストレスレベルが低いと判定された被験者に対して積極的に回答を促す手段が有効であると見込まれる。また、局所バイアスについても、特定の性質を持つ被験者が特定のイベントを経験したタイミングで回答を催促したり、インセンティブを付与したりする設計によって回避できる可能性がある。このような策を講じることで多様な条件下にある数多くの回答データをまんべんなく回収し、ESM データの回答数の最大化や不均一性の解消に貢献することが期待される。ただし、先行研究 [37] によれば、携帯端末の通知による催促は ESM 回答数の増加に総じて有効であるものの、一部の被験者にとっては逆効果をもたらす可能性があるため、どのような手段で催促するかについては慎重に検討する必要がある。

6.4 本研究の制約

本研究は、特定の条件下における ESM データの不均一性を指摘した事例研究である。そのため、本研究の知見は、被験者、イベント、実験設計、環境などが類似した状況、特に被験者が属するコミュニティにおいて影響力のあるイベント^{*3}を経験するような場合 (たとえば、社会人の昇進面談、商談、会合、など) においては有効であると考えられる。

^{*3} あるコミュニティにおける人々の心理や行動に変化をもたらす出来事。本実験においては中間試験や学校祭を指し、より一般的には感染症によるロックダウンや国際的なスポーツイベントなどを含む。

本研究は、同一のデータセットを用いた従来研究 [11] をもとに適合度指標を採用したが、我々の LGC モデルの適合度は十分に高いとはいえない結果となった。ただし、これは先行文献 [42] が指摘しているように、サンプル数の少なさに依存する問題だと考えられる。たとえば、RMSEA や TLI はサンプル数 250 以上が推奨されているが、本研究のサンプル数は 42 である（従来研究 [11] は 33 から 46 である）。そのため、RMSEA では十分なモデル適合度 (0.10 未満 [4], [5], [29]) が確認されなかった。一方、CFI や TLI については一定水準 (0.90 以上 [2]) の結果が示された。これらを総合的に考慮すると、RMSEA の基準が満たされていない点についてはサンプル数の影響を受けたのではないかと考えられる。ただし、統計的に支持されていない可能性も残されているため、将来的に複数のデータセットを統合的に利用してサンプルサイズを拡充し検証する必要がある。

サンプルサイズが小さい点は本研究の重要な制約の 1 つである。ただし、ESM を採用した研究において個人特性と回答数の時間変化の対応関係が記録されており、かつその関係性の調査に利用可能なオープンデータセットは、我々の知る限り StudentLife データセット以外に存在しない。そのため、複数のデータセットを統合的に扱うメタ分析などによってサンプルサイズを拡大させることは現状困難であり、今後の課題として新たなデータセットの構築に取り組むことが求められる。一方で、本研究の実証分析で示された結果は ESM 回答数の個人内時間変動を見たという点で従来研究にない新規的な観点を含んでおり、一事例として報告することが当該研究領域に対する貢献になると我々は期待している。

StudentLife プロジェクトは米国に住む大学生を対象に実施されている。また、本分析からは ESM 回答傾向には定期試験や学校祭といったイベントの影響も確認されている。そのため、異なる文化圏の人々を対象にした場合に本研究の知見の適用可能性は制約的であると考えられる。将来的には、国籍や地域といった地理的特性から形成される生活様式や、大学内もしくは国内における特定のイベントに対する人々の認識の影響について検証する必要がある。これは前述のサンプルサイズの問題と同様に、メタ分析を採用することが有効な手段になると思われる。

我々は、中間試験や学校祭といった、対象となる被験者の多くが共通して経験するイベントを考慮しているが、より個人的な経験については分析していない。たとえば、被験者は家庭内トラブルや交通事故のようなネガティブな経験をするとその被験者の主観的ストレスレベルや内向性に関係なく ESM 質問に回答することをやめてしまう可能性がある。我々の分析は、個人特性と時間の双方に焦点を当てており、上記のような個人的経験の影響を検証できていない。

また、StudentLife プロジェクトにおいて、被験者は活動量リストバンドやスマートタブレットといった報酬を受け取っている。3.2 節で論じたように、本分析で抽出した 42 名の被験者は、上記の報酬によって実験参加への動機付けがなされていたと考えられる。しかし、今後、新たな実験の実施や別のデータセットの利用などによってサンプルサイズを増やす場合、追加されるデータの被験者が ESM 回答に対して十分なモチベーションを持っていなかった可能性がある。そのため、将来研究においてデータを拡充する際には、3.2 節で実施した手続きと同様に、極端に回答数の少ない被験者を除外することで、被験者の ESM 回答に対するモチベーションの高さを統制しなければならない点に留意が必要である。

さらに、本分析で示された結果は StudentLife データセットに依存した一事例である。ただし、先述のように、ESM 研究は研究結果の頑健性について理解を深めるために収集データの偏りについて報告し、研究コミュニティに共有すべきであると我々は考えている。本研究はこの点において確かな貢献を示しているといえる。

6.5 今後の展望

被験者の個人的な経験が ESM 回答数に及ぼす影響をとらえるために、携帯端末で取得された移動、会話、運動に関するログを分析することが有効である。なぜなら、このようなログを複合的に利用することで被験者の経験を抽出できる可能性が高いからである。携帯端末で記録されたログと ESM 回答数の関係を調べ、リアルタイムで経験している状況が ESM 回答数の個人差をどの程度説明するかの検証は今後の検討課題とする。

より一般的に ESM データがどのように偏るかを理解するためには、各要因を統制した様々な条件下で個別に実験を行い、それぞれの実験データを集約したメタ分析を行う必要がある。たとえば、異なる文化圏（東洋か西洋か）や社会的地位（社会人か学生か）を持つ被験者を対象とする実験が想定される。それらの結果との比較や本研究の実験データに基づく統合的な解析により、汎用性のある知見が得られることが期待される。

7. 結論

ESM データは不均質である。しかし現在では、人々の主観的な状態の時間的な変化を追跡する真のデータとして認められ、活用されている。したがって、ESM データにどのような偏りが存在するかを議論することは重要である。本研究は、これまでに ESM データの偏りを調べてきた先行研究とは異なり、ESM 回答数の被験者内時間変化を分析し、どのような被験者の ESM 回答がどのようなタイミングで多いか（増えるか）もしくは少ないか（減るか）を明らかにした。

StudentLife データセットを分析して得られた主な結果として、主観的ストレスレベルの高い被験者は実験期間全体の回答数は多いが、時間の経過にともなう減少傾向が強くなり、特に中間試験終了後に回答頻度が大きく落ち込むことが分かった。また、内向的な被験者は、時間の経過にともなう回答数の減少傾向は弱いですが、学校祭の後に回答頻度が強く減少する傾向を持つことが確認された。さらに、誠実性の高い被験者は、実験の初期段階では回答数が多いが、中間試験によって回答数が減少することから全体として回答頻度の減少傾向が強かった。調和性の高い被験者は、実験開始後に最初に ESM 回答するまでが遅いということが確認された。これらの結果に基づいて、ESM データの個人特性と時間による偏りについて整理し、ESM 研究における根本的なデータの不均一性やその対応における推奨事項を議論した。

今後は、より個人的な経験がもたらす ESM データの偏りを理解するために個人毎の行動や環境に関するセンサーログを活用し、より一般的な知見を導き出すために複数のデータセットを用いたメタ分析を実施する予定である。本研究が、今後の ESM 研究の一助になることを期待する。

参考文献

- [1] Bentler, P.M.: Comparative Fit Indexes in Structural Models, *Psychological Bulletin*, Vol.107, No.2, pp.238–246 (1990).
- [2] Bentler, P.M. and Bonett, D.G.: Significance Tests and Goodness of Fit in the Analysis of Covariance Structures, *Psychological Bulletin*, Vol.88, No.3, p.588 (1980).
- [3] Boukhechba, M., Cai, L., Chow, P.I., Fua, K., Gerber, M.S., Teachman, B.A. and Barnes, L.E.: Contextual Analysis to Understand Compliance with Smartphone-based Ecological Momentary Assessment, *Proc. 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*, pp.232–238, ACM (2018).
- [4] Browne, M.W. and Cudeck, R.: Alternative Ways of Assessing Model Fit, *Sociological Methods & Research*, Vol.21, No.2, pp.230–258 (online), DOI: 10.1177/0049124192021002005 (1992).
- [5] Browne, M.W. and Mels, G.: *RAMONA user's guide* (1992).
- [6] Cohen, S., Kamarck, T. and Mermelstein, R.: A Global Measure of Perceived Stress, *Journal of Health and Social Behavior*, Vol.24, No.4, p.385 (1983).
- [7] Costa, P.T. and MacCrae, R.R.: *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) Manual*, Psychological Assessment Resources, Incorporated (1992).
- [8] Dejonckheere, E., Mestdagh, M., Houben, M., Rutten, I., Sels, L., Kuppens, P. and Tuerlinckx, F.: Complex affect dynamics add limited information to the prediction of psychological well-being, *Nature Human Behaviour*, Vol.3, No.5, pp.478–491 (2019).
- [9] Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-w., Oishi, S. and Biswas-Diener, R.: New Well-being Measures: Short Scales to Assess Flourishing and Positive and Negative Feelings, *Social Indicators Research*, Vol.97, No.2, pp.143–156 (2010).
- [10] Eagle, N., Pentland, A. and Lazer, D.: Inferring friendship network structure by using mobile phone data, *Proc. National Academy of Sciences*, Vol.106, No.36, pp.15274–15278 (2009).
- [11] Harari, G.M., Gosling, S.D., Wang, R., Chen, F., Chen, Z. and Campbell, A.T.: Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods, *Computers in Human Behavior*, Vol.67, pp.129–138 (2017).
- [12] Helgoe, L.A.: *Introvert Power: Why Your Inner Life is Your Hidden Strength*, Sourcebooks, Inc. (2013).
- [13] Hsieh, H.-P. and Li, C.-T.: Inferring social relationships from mobile sensor data, *Proc. 23rd International Conference on World Wide Web*, pp.293–294, ACM (2014).
- [14] Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H.A., Wen, C.K.F. and Field, M.: Compliance with ecological momentary assessment protocols in substance users: A meta-analysis, *Addiction*, Vol.114, No.4, pp.609–619 (2019).
- [15] Kroenke, K., Spitzer, R.L. and Williams, J.B.W.: The PHQ-9, *Journal of General Internal Medicine*, Vol.16, No.9, pp.606–613 (2001).
- [16] Lane, N., Mohammad, M., Lin, M., Yang, X., Lu, H., Ali, S., Doryab, A., Berke, E., Choudhury, T. and Campbell, A.: BeWell: A Smartphone Application to Monitor, Model and Promote Wellbeing, *Proc. 5th International ICST Conference on Pervasive Computing Technologies for Healthcare*, IEEE (2011).
- [17] Lathia, N., Rachuri, K.K., Mascolo, C. and Rentfrow, P.J.: Contextual Dissonance: Design Bias in Sensor-Based Experience Sampling Methods, *Proc. 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp.183–192, ACM (2013).
- [18] Lucas, G.M., Rizzo, A., Gratch, J., Scherer, S., Stratou, G., Boberg, J. and Morency, L.-P.: Reporting Mental Health Symptoms: Breaking Down Barriers to Care with Virtual Human Interviewers, *Frontiers in Robotics and AI*, Vol.4, No.51, pp.1–9 (2017).
- [19] Messiah, A., Grondin, O. and Encrenaz, G.: Factors associated with missing data in an experience sampling investigation of substance use determinants, *Drug and Alcohol Dependence*, Vol.114, No.2-3, pp.153–158 (2011).
- [20] Morren, M., Dulmen, S., Ouwerkerk, J. and Bensing, J.: Compliance with momentary pain measurement using electronic diaries: A systematic review, *European Journal of Pain*, Vol.13, No.4, pp.354–365 (2009).
- [21] Mount, M.K. and Barrick, M.R.: The Big Five personality dimensions: Implications for research and practice in human resources management, *Research in Personnel and Human Resources Management*, Vol.13, No.3, pp.153–200 (1995).
- [22] Rabbi, M., Ali, S., Choudhury, T. and Berke, E.: Passive and In-Situ assessment of mental and physical well-being using mobile sensors, *Proc. 13th International Conference on Ubiquitous Computing*, pp.385–394, ACM (2011).
- [23] Rintala, A., Wampers, M., Myin-Germeys, I. and Viechtbauer, W.: Response compliance and predictors thereof in studies using the experience sampling method, *Psychological Assessment*, Vol.31, No.2, pp.226–235 (2019).
- [24] Russell, D.W.: UCLA Loneliness Scale (Version 3): Reliability, Validity, and Factor Structure, *Journal of Personality Assessment*, Vol.66, No.1, pp.20–40 (1996).

[25] Sadri, A., Salim, F.D., Ren, Y., Shao, W., Krumm, J.C. and Mascolo, C.: What Will You Do for the Rest of the Day? An Approach to Continuous Trajectory Prediction, *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol.2, No.4, pp.1-26 (2018).

[26] Servia-Rodríguez, S., Rachuri, K.K., Mascolo, C., Rentfrow, P.J., Lathia, N. and Sandstrom, G.M.: Mobile Sensing at the Service of Mental Well-being: A Large-scale Longitudinal Study, *Proc. 26th International Conference on World Wide Web*, pp.103-112 (2017).

[27] Sokolovsky, A.W., Mermelstein, R.J. and Hedeker, D.: Factors predicting compliance to ecological momentary assessment among adolescent smokers, *Nicotine and Tobacco Research*, Vol.16, No.3, pp.351-358 (2014).

[28] Steiger, J.H.: Statistically based test for the number of common factors, *The Annual Meeting of the Psychometric Society* (1980).

[29] Steiger, J.H.: EzPATH: A Supplementary Module for SYSTAT and SYSGRAPH, *Evanston, IL: SYSTAT, Inc.* (1989).

[30] Thewissen, V., Bentall, R.P., Oorschot, M., à Campo, J., van Lierop, T., van Os, J. and Myin-Germeys, I.: Emotions, self-esteem, and paranoid episodes: An experience sampling study, *British Journal of Clinical Psychology*, Vol.50, No.2, pp.178-195 (2011).

[31] Thompson, E.R.: Development and Validation of an International English Big-Five Mini-Markers, *Personality and Individual Differences*, Vol.45, No.6, pp.542-548 (2008).

[32] Toegel, G. and Barsoux, J.L.: How to become a better leader, *MIT Sloan Management Review*, Vol.53, No.3, pp.51-60 (2012).

[33] Tominaga, T., Yamamoto, S., Kurashima, T. and Toda, H.: Effects of Personal Characteristics on Temporal Response Patterns in Ecological Momentary Assessments, *IFIP Conference on Human-Computer Interaction (INTERACT)*, pp.3-22 (2021).

[34] Tucker, L.R. and Lewis, C.: A reliability coefficient for maximum likelihood factor analysis, *Psychometrika*, Vol.38, No.1, pp.1-10 (1973).

[35] Udachina, A., Varese, F., Myin-Germeys, I. and Bentall, R.P.: The role of experiential avoidance in paranoid delusions: An experience sampling study, *British Journal of Clinical Psychology*, Vol.53, No.4, pp.422-432 (2014).

[36] Vachon, H., Viechtbauer, W., Rintala, A. and Myin-Germeys, I.: Compliance and retention with the experience sampling method over the continuum of severe mental disorders: Meta-analysis and recommendations, *Journal of Medical Internet Research*, Vol.21, No.12 (2019).

[37] van Berkel, N., Goncalves, J., Hosio, S., Sarsenbayeva, Z., Velloso, E. and Kostakos, V.: Overcoming compliance bias in self-report studies: A cross-study analysis, *International Journal of Human-Computer Studies*, Vol.134, pp.1-12 (2020).

[38] van Berkel, N., Goncalves, J., Lovén, L., Ferreira, D., Hosio, S. and Kostakos, V.: Effect of experience sampling schedules on response rate and recall accuracy of objective self-reports, *International Journal of Human-Computer Studies*, Vol.125, pp.118-128 (2019).

[39] Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D. and Campbell, A.T.: StudentLife: Assessing Mental Health, Academic Performance and Behavioral Trends of College Students using Smartphones, *Proc. 2014 ACM International Joint*

Conference on Pervasive and Ubiquitous Computing, pp.3-14, ACM (2014).

[40] Wang, R., Harari, G., Hao, P., Zhou, X. and Campbell, A.T.: SmartGPA: How Smartphones Can Assess and Predict Academic Performance of College Students, *Proc. 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp.295-306 (2015).

[41] Yao, H., Lian, D., Cao, Y., Wu, Y. and Zhou, T.: Predicting Academic Performance for College Students: A Campus Behavior Perspective, *ACM Tran. Intelligent Systems and Technology*, Vol.10, No.3, pp.1-21 (2019).

[42] 星野崇宏, 岡田謙介, 前田忠彦: 構造方程式モデリングにおける適合度指標とモデル改善について: 展望とシミュレーション研究による新たな発見, *行動計量学*, Vol.32, No.2, pp.209-235 (2005).



富永 登夢

2016年大阪大学大学院基礎工学研究科修士課程修了。2019年同研究科博士課程修了。博士(工学)。2017年米国ミネソタ大学客員研究員。2019年日本電信電話(株)入社。現在, NTT人間情報研究所研究員。計算社会科学, データマイニング, HCIに関する研究開発に従事。



山本 修平 (正会員)

2014年筑波大学大学院図書館情報メディア研究科博士前期課程修了。2016年同研究科博士後期課程修了。博士(情報学)。同年日本電信電話(株)入社。現在, NTT人間情報研究所研究主任。データマイニング, 時空間データ分析に関する研究開発に従事。日本データベース学会会員。



倉島 健 (正会員)

2006年京都大学大学院情報学研究科修士課程修了。2014年同研究科博士課程修了。博士(情報学)。2006年日本電信電話(株)入社。2016~2017年米国スタンフォード大学客員研究員。現在, NTT人間情報研究所主任研究員。人間行動に関するデータマイニングと機械学習技術の研究開発に従事。電子情報通信学会, 日本データベース学会, ACM各会員。



戸田 浩之 (正会員)

1999年名古屋大学大学院工学研究科博士課程前期課程修了。2007年筑波大学大学院システム情報工学研究科博士後期課程修了。博士(工学)。1999年日本電信電話(株)入社。2016~2022年NTTサービスエボリューション研

究所, NTT人間情報研究所主幹研究員。現在, 横浜市立大学データサイエンス学部教授。これまで情報検索, データマイニングの研究開発に従事。電子情報通信学会, 日本データベース学会, 人工知能学会, ACM各会員。