

THE IEICE TRANSACTIONS ON INFORMATION AND SYSTEMS (JAPANESE EDITION)

EIC | **電子情報通信学会** **D** | **論文誌**

情報・システム

VOL. J104-D NO. 5

MAY 2021

本PDFの扱いは、電子情報通信学会著作権規定に従うこと。

なお、本PDFは研究教育目的（非営利）に限り、著者が第三者に直接配布することができる。著者以外からの配布は禁じられている。

情報・システムソサイエティ

一般社団法人 **電子情報通信学会**

THE INFORMATION AND SYSTEMS SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

性別に着目したソースコード理解速度の分析*

高塚由利子^{†a)}村上優佳紗[†]角田 雅照[†] (正員)中村 匡秀^{††} (正員)

A Gender-Focused Analysis of Source Code Reading Speed *

Yuriko TAKATSUKA^{†a)}, Yukasa MURAKAMI[†], *Nonmembers*,Masateru TSUNODA[†], and Masahide NAKAMURA^{††}, *Members*[†] 近畿大学大学院総合理工学研究科, 東大阪市

Graduate School of Science and Engineering, Kindai University, Higashi-Osaka-shi, 577-8502 Japan

^{††} 神戸大学大学院システム情報学研究科, 神戸市

Graduate School of System Informatics, Kobe University, Kobe-shi, 657-8501 Japan

a) E-mail: takayuri0906@gmail.com

* 本論文は高塚 由利子他, “ソフトウェア開発における性別とプログラム読解速度との関係,” ソフトウェア工学の基礎 XXVI, 日本ソフトウェア科学会, 2019 の拡張である.

DOI:10.14923/transinf.2020JDL8013

あらまし 女性の人材を増やすためには, 女性に関する先入観の影響を抑え, 女子学生の進路の選択肢を増やすことなどが挙げられる. そこで本研究では, そのような先入観に対処するために, 性別に着目してプログラム理解速度を分析した. 実験の結果, 性別によるプログラム理解速度の差は見られなかった.

キーワード 開発者の能力, コードリーディング, メンタルシミュレーション

1. まえがき

近年, IT 人材の不足が指摘されている [1]. 約 70% の企業では, 女性の IT 人材が 20% 以下となっており [2], 現状は男性技術者の方が多い. すなわち, 女性 IT 人材を活用する余地があると考えられ, 女性の IT 技術者を増やすことで人材不足の解決の一助となる. 更に, 女性 IT 人材が増えることは, 男性, 女性双方にとってメリットがある. 例えば Hoogendoorn ら [3] の研究では, 性別に偏りがないビジネスチームでは, そうでないチームと比較してより高い利益をあげられることが示されている.

女性の IT 人材を増やすためには, 科学技術が男性に向いているなどの先入観の影響を抑え, 女子学生の進路の選択肢を増やすこと [4] などが重要な要素として挙げられる. このような問題に対処する一つの方法は, 定量的な証拠を示すことである.

そこで本研究では, 上述のような先入観に対処するために, プログラム理解速度と性別との関係について分析する. ソフトウェアライフサイクルコストの 40-67% は保守であり [5], 保守時にプログラム理解が

必要となるため, 理解速度は重要な能力の一つといえる. プログラムの理解には記憶力が必要とされる場合があり, 男性と女性では記憶能力に違いがあると指摘されていることから [6], 性別がプログラム理解速度に影響する可能性がある. 一方で, 3.2 で述べるように記憶能力のプログラム理解への影響度は必ずしも明確でない. そこで, 理解速度が記憶能力に影響されやすいプログラムを読む場合, 性別により理解速度に差があるかを確かめる.

2. 関連研究

本研究と同様に, 女性をサポートする目的で性別に着目して分析した研究は幾つか存在する. 例えば Terrell ら [7] は, 性別に着目して GitHub におけるプルリクエストの採択される割合を分析し, 性別が不明な場合は女性が採択される割合が高く, 性別が明らかな場合は逆の結果となったことを示している. これは, 性別に関する先入観を抑えることにより, 女性 IT 人材を活用できる可能性を示している. また, ソフトウェアのユーザインタフェースを女性が使いやすくするために, 性別によりソフトウェアの操作が異なるかを分析した研究も存在する [8], [9]. 更に, コンピュータサイエンスを専攻する女子学生が減少している原因を明らかにするために, 男子学生と女子学生のコンピュータサイエンスに対する認識の違いを分析した研究 [10] も存在する. この研究は, 学生の支援や大学カリキュラムの改善に役立ち得る.

ただし我々の知る限り, 性別に着目してソースコード理解速度を実験的に分析した研究は存在しない. 一方で, 本研究とこれらの従来研究は, 女性 IT 技術者とその候補をサポートするというゴールは同じである. なお本研究では, 科学技術が男性に向いているなどの先入観の影響を抑え, 女子学生の進路の選択肢を増やすために, 定量的な証拠を示すことを目的としており, その点においては異なる.

3. プログラム理解

3.1 ボトムアップ理解

プログラム理解には, トップダウン理解とボトムアップ理解がある [11]. 前者では意味的な手がかり (semantic cues) に基づいてプログラムを理解する. 例えば bubble sort という識別子がプログラム内に使われている場合, 開発者はバブルソートを行っていると推測して理解を試みる. これに対しボトムアップ理解では, 意味的な手がかりが乏しい場合に適用され, 各プログラム行を解釈し, そこからプログラム全体を理解

するアプローチを取る。

これまで、トップダウン理解を扱っている研究[11]もあるが、ボトムアップ理解に基づくプログラム理解を取り扱っている研究も多く見られる。例えば文献[12]ではプログラム理解時の脳血流を分析するために、被験者にプログラム内の各変数の値を答えさせている。文献[13],[14]でも、プログラム理解時の脳波を分析するために、被験者にプログラムの出力を答えさせている。プログラム保守においても意味的な手がかりが乏しい場合、ボトムアップ理解を行う必要があるため、本研究ではこれらの研究と同様に、ボトムアップ理解を前提とする。すなわち、本研究でのプログラム理解速度とはボトムアップ理解時における理解速度を指す。実験は、被験者がボトムアップ理解する必要のあるプログラムを用いて実施した。

3.2 記憶能力の影響

プログラムの理解には記憶能力が必要とされる場合があるが、プログラムをボトムアップ理解するためには、単に変数の値を記憶しておくだけでは不十分である。理解のためには条件分岐についても考慮する必要があり、条件式に応じて分岐先を選び、時には記憶している変数の値も更新する必要がある。すなわち、プログラム理解は記憶能力以外も求められる複雑なタスクであり、記憶能力の違いがプログラム理解にどこまで影響するのが必ずしも明らかでないといえる。

なお、本研究では、プログラム理解速度に影響する記憶能力は単純な短期記憶ではなく、ワーキングメモリの能力であると想定する。短期記憶は受動的な一時記憶であるのに対し、ワーキングメモリは記憶と処理の両方に関連する能動的な一時的記憶であり、情報を更新したり調整したりする側面をもつ[15]。そのため、ボトムアップ理解の場合、ワーキングメモリの能力差がプログラム理解速度に影響すると考えられる。ただし、ワーキングメモリには個人差があり、また容量にも制約があるため、効率的な運用には注意による制御が欠かせないと指摘されている[15]。そこで本研究では主にワーキングメモリ容量の運用能力を、記憶能力として想定する。

3.3 理解容易性尺度

あるプログラムの理解速度が、記憶能力、すなわちワーキングメモリ容量の運用能力に影響されやすいかを定量的に計測するため、文献[16],[17]で提案されている理解容易性尺度を用いた。この尺度では、プログラムの理解時にメンタルシミュレーション[18]が行わ

れることを想定している。メンタルシミュレーションとは、プログラムの動作をコンピュータや筆記具を用いずに思考しながら理解することであり、比較的規模の小さなコード片に対して用いられる。

メンタルシミュレーションを行う場合、変数の値を記憶しておく必要があるが、多数の値を同時に記憶しておくことは容易ではない。それを考慮した尺度として、文献[17]ではBT_CONST, ASSIGN, RCL, BT_VARが定義され、文献[16]ではBT_VARを改善したSUM_UPD, SUM_VARが定義されている。そこで本研究ではBT_VARを除いた五つの尺度を用いた。BT_CONST, SUM_UPD, SUM_VARでは、変数の値が開発者の一時記憶に存在しない場合に、その値を得るための再計算コストに基づいて定義されている。尺度の値が大きい(SUM_VARについては小さい)プログラムでは、再計算コストが高いことを示す。記憶能力が低く、一時記憶に保持できる変数が少ない開発者がそのようなプログラムを読むと、再計算の頻度が高まり、かつ再計算コストが大きいため、理解速度が低下しやすい。

また、文献[17]で定義されている尺度ASSIGN, RCLは開発者の一時記憶への読み書き回数に基づいている。一時記憶の読み書きに時間が掛かる、すなわち記憶能力が低い場合、これらの尺度の値が大きいと、理解速度が遅くなる。よって、これらの尺度の値が大きい(SUM_VARは小さい)場合、理解速度が記憶能力に影響されやすいといえる。

4. 実験

実験の目的は、理解速度が開発者の記憶能力に影響されやすいプログラムを読む場合、性別により速度が異なるかどうかを確かめることである。そのために、記憶能力の理解速度への影響度が異なる、複数のプログラムを用意し、男女の被験者がコードを理解するために掛かった時間を計測した。開発者の能力には様々な側面が存在するが、本実験では総合的な能力の評価を目的とせず、特に記憶能力のみに着目している。

実験では文献[16]で示されているプログラム四つ(a0, a1, b0, b1)を利用した。それぞれ20行から30行の規模である。ある変数の値が、プログラム実行後にどうなるかを正しく答えられる時間を回答時間、すなわち理解速度とした。より詳細には、回答の開始はプログラムを被験者に提示した時点とし、終了は正答を答えた時点とした。表2に各プログラムの理解容易性尺度(3.参照)を示す。これらより、b0, b1の理解

表1 被験者に関する基本統計量
Table 1 Basic statistics of subjects.

		Avg.	Med.	S.D.
男性	年齢	21.3	21.0	0.5
	経験年数	4.3	3.0	2.2
女性	年齢	22.2	22.0	1.5
	経験年数	3.0	3.0	0.8

表2 各プログラムの理解容易性尺度 [16]
Table 2 Program understandability metrics on each program.

	ASSIGN	RCL	BT_CONST	SUM_UPD	VAR_UPD
a0	12	6	0	7	1.25
a1	12	6	0	7	0.25
b0	18	8	1	11	0.24
b1	18	6	1	11	0.24

速度は記憶能力に影響されやすいといえる。なお、文献 [16] の表では BT_VAR が含まれていたが、3.3 で述べたように BT_VAR は用いないため、表から除外した。

被験者を男性、女性グループに分け、それぞれのグループの回答時間の平均値や中央値などを算出し比較した。被験者は、同一大学・学科の情報科学を専攻する学部生 16 名（男性 8 名、女性 8 名）である。被験者の経験年数などの基本統計量を表 1 に示す。男性グループの経験年数の平均が 1 年長いが、これは経験年数が 7 年以上の被験者が 2 人含まれるためであり、それ以外の被験者の経験年数は、男女とも 2~4 年であった。

プログラムを読む順番が実験結果に影響することを避けるために、四つのプログラムを読む順番を被験者ごとに変更した。例えばある被験者ではプログラムを a0, a1, b0, b1 の順に読むとし、別の被験者では b1, a0, b0, a1 の順に読むなどとした。

分析の目的を明確にするために、三つのリサーチクエスションを設定した。RQ1, RQ2 はグループ間の速度差に着目しているのに対し、RQ3 はグループ内における、プログラムの違いによる速度差に着目している。

- RQ1: 女性グループと男性グループでは、どちらがプログラムを理解する速度が速いのか?
- RQ2: プログラムの理解速度が開発者の記憶能力に影響されやすい場合、どちらのグループの理解速度が速いのか?
- RQ3: プログラムの理解速度が開発者の記憶能力に影響されやすい場合とそうでない場合の理解速度の差が小さいのは、どちらのグループか?

表3 各グループのプログラム別回答時間 (秒)
Table 3 Correlation coefficients to reading time.

		a0	a1	b0	b1
男性	平均値	51.4	93.6	90.5	156.4
	中央値	38.5	89.5	71.0	87.0
	標準偏差	25.5	48.2	55.4	141.4
女性	平均値	57.4	98.5	160.6	120.1
	中央値	54.5	86.0	133.5	110.0
	標準偏差	18.8	58.3	114.7	37.6

5. 結 果

5.1 RQ1, RQ2 に関する分析

表 3 に男性、女性各グループのプログラム別の回答時間の平均値、中央値、標準偏差を示す。時間が短いグループのセルをグレーで表している。プログラム a0, b0 では、平均値、中央値とも男性グループの方が小さかったが、プログラム a1 では女性グループの中央値が小さく、b1 では平均値が女性グループの方が小さかった。すなわち、プログラムによって結果が異なり、必ずしも一方のグループの時間が短いとはいえない。

それぞれのグループの回答時間の分布を、箱ひげ図を用いて図 1 に示す。a0 では箱の位置は両グループで差がないが、男性グループの中央値の方が低い。a1 では箱の位置、中央値ともほとんど差がない。b0 では男性グループの箱の位置、中央値とも低かった。b1 では中央値は若干男性の方が低かったが、箱の大きさは男性の方が大きかった。図からもどちらかのグループが常に時間が短いとはいえない。

この結果より、RQ1 に対する答えは「プログラムによって傾向が異なり、必ずしも一方が速いとはいえない」、RQ2 に対する答えは「プログラムによって傾向が異なり、どちらかのグループの方が速いとはいえない」となる。

5.2 RQ3 に関する分析

RQ3 に答えるために、理解速度に対する記憶能力の影響が小さいプログラム a0 と比べ、何倍の回答時間が掛かっているかを調べた。具体的には、プログラム a0 以外の回答時間/プログラム a0 の回答時間を求めた。この値が小さければ、記憶能力に影響されやすい場合とそうでない場合の理解速度の差が小さいといえる。なお a0 の回答時間を基準とすることにより個人の能力差を標準化している、言い換えると a0 を各個人の基準値とし、基準値との差を分析しているともいえる。

結果を表 4 に示す。プログラム a1, b1 では、女性グループの方が比の平均値、中央値とも小さかったが、

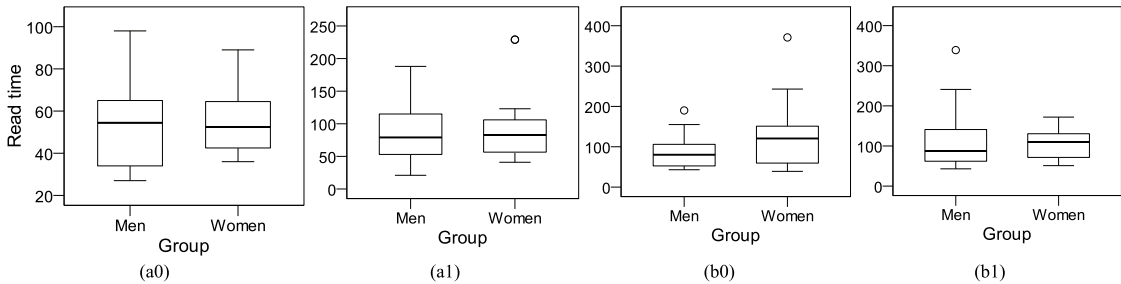


図1 プログラム別回答時間 (秒)
Fig. 1 Answering time of each program (seconds).

表4 プログラム a0 との回答時間の比
Table 4 Ratio of answering time to program a0.

		a1 / a0	b0 / a0	b1 / a0
男性	平均値	1.83	1.68	2.61
	中央値	1.97	1.83	2.29
	標準偏差	0.45	0.31	1.03
女性	平均値	1.63	2.51	2.11
	中央値	1.52	2.47	2.10
	標準偏差	0.40	1.11	0.17

表5 誤回答数の基本統計量
Table 5 Basic statistics of the number of incorrect answers.

		a0	a1	b0	b1
男性	平均値	0.38	0.50	0.38	1.88
	中央値	0.00	0.00	0.00	0.50
	標準偏差	0.74	1.07	0.74	3.00
女性	平均値	0.38	1.13	1.63	1.13
	中央値	0.00	1.00	1.00	1.00
	標準偏差	0.52	1.64	1.85	1.36

プログラム b0 では男性グループの方が小さかった。すなわち、回答時間の相対的な変化量に着目した場合でも、プログラムによって結果が異なり、どちらかのグループの時間が短いとはいえなかった。

5.3 誤回答数に関する分析

誤回答数の基本統計量を表5に示す。プログラム b0 については男性グループの誤回答数が多く、b1 では女性グループの誤回答数が多かったが、これらはそれぞれのグループで回答時間が長いプログラムであり、難易度がそのまま反映されていると考えられる。女性グループでは、a0 を除き中央値が1を超えていた。このことから、誤回答の傾向は両グループで特に大きな差はないが、少なくとも女性グループの誤回答数が少ない傾向とはいえない。

6. 考 察

6.1 妥当性への脅威

実験での被験者は同一大学・学科に所属しているため、受けてきたプログラミング教育はおおむね同じと

いえる。また、プログラミング経験年数の違いは大きくなく、更に個人ごとに回答時間を正規化した分析も行った。このことから、性別以外の要因については適切にコントロールできていると考える。

本研究では、性別以外の要因として、表1に示すように年齢と経験年数を確かめており、従来研究と比較して要因の考慮は不足していない。文献[14]では学生の被験者は比較的均一であるとみなし、そもそも経験年数などの要因が確かめられていない。文献[19]ではプログラミング言語ごとの経験年数のみ確かめており、文献[9]は年齢、経験年数と4段階の技術レベルを確かめている。学生の自己評価による技術レベルは、我々のこれまでの実験において、実際の技術レベルと関連が弱い傾向が見られたため収集しなかった。

実験では規模の小さなプログラムを用いた。もし、より規模の大きなプログラムを用いた場合、理解に必要な知識など、その他の要因の影響が大きくなり、記憶力を必要とするプログラムとそうでないものを読んだ場合の差異の要因が不明確となり得る。実験での規模が小さいことは留意すべきだが、文献[20]など、被験者実験時にコードスニペットとも呼ばれる規模の小さなプログラムを用いた研究は幾つか存在する。

実験での被験者はソフトウェア開発の実務者ではない。ただし、文献[19]では、学生と実務者の違いはあまり大きくなく、学生を実務者の代替できることが示されている。実験で用いたプログラムの内容も考慮すると、実務者を被験者とした場合でも分析結果は大きく変わらないと考えられる。ただし、先入観の影響を抑え、女子学生の進路の選択肢を増やすことが本研究の目的であることから、男女の学生間だけでなく、男女の実務者間でも差がないことを定量的に示すことも重要であると考えられる。そのため、実務者を被験者とするのは今後の課題となり得る。

被験者数は多くはないが、被験者実験は実施に時間が掛かり、人数を増やすことは容易ではない。例えば、文献[14]では被験者数が14人である。被験者数を増やして信頼性を高めることは今後の課題の一つである。

6.2 結果の解釈と活用

同一大学・学科の被験者による実験では、性別と理解速度には関連が見られなかった。このことから、プログラミングの経験と教育水準、例えば教育を受けている大学が同等であるならば、対象プログラムの記憶力の必要性にかかわらず、性別によるコード理解速度のアドバンテージはほとんどないと考えられる。

分析結果の活用法の一つは、1.で述べたように先入観の影響を抑えて女子学生の進路の選択肢を増やすことである。更に、先入観の影響を抑えた場合、より直接的な利益を得られる可能性がある。文献[21]では、性別に基づく先入観の悪影響を「ステレオタイプ脅威」と呼んでおり、この脅威を取り除くと、女性の数学のテスト成績が向上することが示されている。分析結果を活用してプログラミングに対するステレオタイプ脅威を取り除くと、女性がプログラミング能力を更に発揮できる可能性がある。

7. むすび

本研究では、性別による先入観を抑え、女性IT人材の増加をゴールとして、性別に着目してソースコード理解速度を分析した。分析の結果、メンタルシミュレーションによりソースコードを理解する場合、性別と理解速度と関連は見られなかった。今後の予定は、被験者を増やし結果の信頼性を高めることである。

文 献

- [1] 総務省, 平成 29 年版 情報通信白書, 総務省, 2017.
- [2] 情報処理推進機構, IT 人材白書 2018, IPA, 2018.
- [3] S. Hoogendoorn, H. Oosterbeek, and M. Praag, "The impact of gender diversity on the performance of business teams: Evidence from a field experiment," *Management Science*, vol.59, no.7, pp.1514–1528, 2013.
- [4] V. Borsotti, "Barriers to gender diversity in software development education: actionable insights from a danish case study," *Proc. Int. Conf. Softw. Eng.: Software Engineering Education and Training (ICSE-SEET)*, pp.146–152, 2018.
- [5] M. Carr and C. Wagner, "A study of reasoning processes in software maintenance Management," *Information Technology and Management*, vol.3, pp.181–203, 2002.
- [6] P. Loprinzi and E. Frith, "The role of sex in memory function: Considerations and recommendations in the context of exercise," *J. Clinical Medicine*, vol.7, no.6, article 132, 2018.
- [7] J. Terrell, A. Kofink, J. Middleton, C. Rinear, E. Murphy-Hill, C. Parnin, and J. Stallings, "Gender differences and bias in open source: pull request acceptance of women versus men," *Peer J Computer Science*, 3:e111, 2017.
- [8] L. Beckwith, M. Burnett, S. Wiedenbeck, C. Cook, S. Sorte, and M. Hastings, "Effectiveness of end-user debugging software features: are there gender issues?" *Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI)*, pp.869–878, 2005.
- [9] M. Burnett, S. Fleming, S. Iqbal, G. Venolia, V. Rajaram, U. Farooq, V. Grigoreanu, and M. Czerwinski, "Gender differences and programming environments: across programming populations," *Proc. Int. Symp. Empirical Software Engineering and Measurement (ESEM)*, article 28, 10 pages, 2010.
- [10] S. Beyer, K. Rynes, J. Perrault, K. Hay, and S. Haller, "Gender differences in computer science students," *Proc. 34th SIGCSE technical symposium on Computer science education (SIGCSE)*, pp.49–53, 2003.
- [11] J. Siegmund, N. Peitek, C. Parnin, S. Apel, J. Hofmeister, C. Kästner, A. Begel, A. Bethmann, and A. Brechmann, "Measuring neural efficiency of program comprehension," *Proc. Joint Meeting on Foundations of Software Engineering (ESEC/FSE)*, pp.140–150, 2017.
- [12] T. Nakagawa, Y. Kamei, H. Uwano, A. Monden, K. Matsumoto, and D. German, "Quantifying programmers' mental workload during program comprehension based on cerebral blood flow measurement: a controlled experiment," *Companion Proc. Int. Conf. Softw. Eng. (ICSE Companion)*, pp.448–451, 2014.
- [13] I. Crk, T. Kluthe, and A. Stefik, "Understanding programming expertise: An empirical study of phasic brain wave changes," *ACM Trans. Computer-Human Interaction*, vol.23, no.1, article 2, 2016.
- [14] J. Siegmund, C. Kästner, S. Apel, C. Parnin, A. Bethmann, T. Leich, G. Saake, and A. Brechmann, "Understanding understanding source code with functional magnetic resonance imaging," *Proc. Int. Conf. Softw. Eng. (ICSE)*, pp.378–389, 2014.
- [15] 荻阪直行, "前頭前野とワーキングメモリ," 高次脳機能研究, vol.32, no.1, pp.7–14, 2012.
- [16] 石黒誉久, 井垣 宏, 中村匡秀, 門田暁人, 松本健一, "変数更新の回数と分散に基づくプログラムのメンタルシミュレーションコスト評価," 信学技報, SS2004-32, 2004.
- [17] M. Nakamura, A. Monden, H. Satoh, T. Itoh, K. Matsumoto, and Y. Kanzaki, "Queue-based cost evaluation of mental simulation process in program comprehension," *Proc. International Software Metrics Symposium*, pp.351–360, 2003.
- [18] A. Dunsmore, M. Roper, and M. Wood, "The role of comprehension in software inspection," *J. Systems and Software*, vol.52, no.2-3, pp.121–129, 2000.
- [19] I. Salman, A. Misirli, and N. Juristo, "Are students representatives of professionals in software engineering experiments?" *Proc. Int. Conf. Softw. Eng. (ICSE)*, pp.666–676, 2015.
- [20] B. Floyd, T. Santander, and W. Weimer, "Decoding the representation of code in the brain: an fMRI study of code review and expertise," *Proc. Int. Conf. Softw. Eng. (ICSE)*, pp.175–186, 2017.
- [21] S. Spencer, C. Steele, and D. Quinn, "Stereotype threat and women's math performance," *J. Experimental Social Psychology*, vol.35, no.1, pp.4–28, 1999.

(2020 年 7 月 23 日受付, 11 月 9 日再受付,
2021 年 2 月 3 日早期公開)