

A plausible connection: Models examining the relations between evaluation, plausibility, and the refutation text effect

Doug Lombardi, Robert W. Danielson, Neil Young, (2016) *Learning and Instruction*, 44, 74-86.

1. Introduction

- 近年の教育は、科学に対するリテラシーや理解の促進を目指している
 - 科学的説明の構築もその一つ
- 科学者は、「与えられた状況において起こりえると予測される観察事象に対して妥当な説明」を生成する
 - 科学的モデルや理論を形成するとき、説明の妥当性を比較する
 - 特定の議題や複雑な話題において、科学者は、学生にとって妥当とは思えない説明を生成し、非科学的説明の方が、科学的説明より妥当に見えることもある
- 妥当性：自動的に起こる認知評価によって、しばしば暗黙のうちに行われる一時的な判断 (Rescher, 1976; Hinze, Slaten, Horton, Jenkins, & Rapp, 2014)
 - 説明に対する妥当性の判断は、批判的思考や知識、概念変化と関連がある
 - 最近の研究では、説明に対する評価と妥当性の関連を示す
- 代替説明の評価は、批判的思考の重要な側面の一つ (West, Toplak, & Stanovich, 2008)
 - 教育では、学生が現象の説明や問題の意見に対する評価を、より批判的に考えることを支援する (McCrudden & Sparks, 2014; Nussbaum, 2008; Muis & Duffy, 2013)
- 学生の誤った概念を変えることは難しい (Chi, 2005)
 - 反証的文章は、特定の状況下での概念変化を促進する教育的手法の一つ (Sinatra & Broughton, 2011)
 - ◇ 反証的文章：既存の誤った知識を述べ、直接論破し、正しい知識を提供する文章 (Kendeou, Walsh, Smith, & O'Brien, 2014)
 - ◇ 対になる議論を提示し、認知的葛藤を促進することで、概念変化を促す (Guzzetti, Snyder, Glass, & Gamas, 1993; Hynd, 2001)
 - ◇ 反証的文章は、様々なトピックで知識の再構築を促進することにおいて、他の文章構造に勝る (Diakidoy, Kendeou, & Ioannides, 2003; Broughton, Sinatra, & Reynolds, 2010; Cordova, Sinatra, Broughton, Taasoobshirazi, & Lombardi, 2014)
 - 文章は、長期記憶の情報の統合化ネットワークを構築することで、新規の概念を強化する
 - ◇ 反証的文章は概念変化を促進し、同時に、既存の概念による干渉を減少させるかもしれない (Kendeou & O'Brien, 2014)
 - 概念変化が行われるには、新規の概念の妥当性が、既存の概念の妥当性より優れていなくてはならない (Lombardi et al., 2015; Posner et al., 1982)

- 概念変化における妥当性判断モデル(Lombardi et al., 2015) (Fig. 1)
 - もし新規の概念が既存の概念より妥当であるなら、概念変化が起こる
 - ◇ 妥当性の再評価は、知識の再構築の構成要素の可能性
 - 本研究の目的は、ロンバルディら(2015)の理論的位置づけを実験的に検討する

The present study

- 代替説明の評価と妥当性の再評価の間の理論的な関連を提唱 (Lombardi et al., 2015)
 - この潜在的関係を直接検討した実験的な研究はされていない
 - 説明の妥当性の再評価と批判的な評価をする能力を駆動させるために反証的文章を使用した実験的研究も見当たらない
- 本研究の目的
 - 批判的な評価をする能力と妥当性の再評価の間に重要な関連があるという理論的主張の実験的検討
- リサーチクエスチョン①
 - 反証的文章を読んだときの評価と妥当性の潜在的関連を調査
 - ◇ 批判的に評価する能力、妥当性についての判断、標準的な説明的文章か反証的文章のどちらか一方を読むことに基づいた異なる気候変動の知識の間にどのような関連があるか
 - ◇ 仮説 1a : 反証的文章を読むことが批判的な評価能力を活性化させ、人が現在の気候変動を引き起こしているという科学的に認められている意見へ妥当性をシフトし、気候変動についての知識を増やす
 - 人が新しい情報に直面するとき、説明に対する評価が妥当性の再評価を引き起こすかもしれない
 - ◇ 仮説 1b : 評価、妥当性、説明的文章による知識の間に有意な関連性はない
 - 分かりやすく明確な説明は、必ずしも批判的思考や、妥当性の再評価、知識の再構築を促進しない (Dole & Sinatra, 1998; Lombardi et al., 2015)
 - 2種類の文章の変数間の関連性を示す仮定的モデルを開発 (Fig. 2)
- リサーチクエスチョン②
 - 気候変動についての反証的文章が、妥当性と知識をどのように変化するか
 - ◇ 仮説 2a : 反証的文章を読んだ参加者は、標準的な説明的文章を読んだ参加者より、より大きく妥当性をシフトする
 - ◇ 仮説 2b : 反証的文章を読んだ参加者は、標準的な説明的文章を読んだ参加者より、気候変動についての知識が変わる

2. Methods

- 参加者
 - 大学生 95 名 ($M_{age} = 19.40$, $SD_{age} = 1.78$)
- 課題

1. 提示文章

1.1. 説明的文章

- ◇ 様々なプリントとインターネットソースを組み合わせて作成した（811 words）
- ◇ 地球が気温を維持し、管理する自然なプロセスを持っていることが主な内容

- Ex.

「地球の温室効果は地球の平均気温を上げることを助け、生物が居住可能な状態にする自然な事象です。水蒸気や二酸化炭素のような自然に生じる温室効果ガスなしでは、多くの地球のエネルギーは宇宙空間へと放出し、地球の平均気温は-1ℱ (-18℃) 程度になり、今日よりおよそ 60ℱ (16℃) 寒くなります。地球での生活は、温室効果なしでは非常にことなるでしょう」

- ◇ さらに、人間が温室効果を強化し、世界的な気温上昇につながったことについても議論している

- Ex.

「自然に発生する温室効果は地球を居住可能な状態に維持しています。しかし、強化された温室効果は、現在の地球温暖化の原因となります。科学的な観察は、世界的に平均気温が上昇していることを確認しています。大部分の温暖化が1970年代から起こり、表面の気温は1880年以降増加しました…」

1.2. 反証的文章

- ◇ 説明的文章の情報や文章量をベースに、いくつかの反論意見を加えた（999 words）

- Ex.

「自然な地球温暖化が『温室効果』と呼ばれると、一部の人々は、物理的にエネルギーを閉じ込める、地球を取り囲んでいる大気にチリやガスの薄い層があるという、誤った想像をしてしまいます。しかし、これは正しくありません。地球の周囲の大気に、物理的に熱を閉じ込めるチリやガスの薄い層はありません。対照的に、地球の大気に入ってくるエネルギーは、宇宙空間に出ていくエネルギーと釣り合っています。人々は、自然に発生する温室効果を世界的な平均気温の上昇と混同し、しばしば、地球温暖化と呼びます。一部の人は地球温暖化が起きていることを信じません。あなたは、また、世界的に気温が上昇していることを疑ったかもしれません。それでも、世界的に平均気温が上昇していないと思うことは、誤りです。大部分の温暖化が1970年代から起こり、表面の気温は1880年以降増加しました…」

2. 気候変動に対する知識測定 (HICCK; Lombardi et al., 2013a)

- ◇ the Human-Induced Climate Change Knowledge instrument
- ◇ 人間によって引き起こされた気候変動における現在の科学的意見についての理解を測定 (27項目, $\alpha = .79$)
 - 地球温暖化に関連した科学的理解や人間による気候変動についての日常的な概念、累積的研究が基盤

- ◇ 気候科学者が、それぞれの意見を述べるかどうか (5 段階評定)
 - Ex. 「現在の気候変動は太陽エネルギーの増加によって引き起こされる」
- 3. 妥当性評定
 - ◇ 現在の気候変動を説明する 2 つのモデルに対してもっともらしさを測定する
 - Model A : 人間によって引き起こされた気候変動 (科学的に認められたモデル; Doran & Zimmerman, 2009)
 - Model B : 太陽エネルギーの増加 (有名な懐疑論者のモデル; Lombardi et al., 2013a)
 - ◇ 参加者は両方のモデルの妥当性を「非常に信じられない(greatly implausible; 1)」～「非常に信じられる(highly plausible; 10)」の 10 段階で評定
 - ◇ 妥当性得点 : Model A の評価から Model B の評価を引いたもの
 - ポジティブな妥当性評価 (+) : 参加者が非科学的な説明よりも科学的な説明の妥当性に気づく
 - ネガティブな妥当性評価 (-) : 非科学的説明がよりもっともらしく見える
 - どちらでもない (0) : どちらの説明ももっともらしく見える
- 4. 批判的評価課題 (Critical Evaluation Task: CET)
 - ◇ 欠陥を含んだ文章を読んで欠陥を発見する課題 (Glassner and Schwarz, 2005)
 - 「欠陥を見つける」という手法は、批判的思考に関する推論手法の一つ (Sternberg, 2001)
 - 現象に対する代替説明と証拠の結びつきに重みをつける科学的プロセスに似ている (McNeill, Lizotte, Krajcik, & Marx, 2006)
 - ◇ 参加者は、死刑についての説明と死刑に反対する説明に対して、いくつかの欠陥を考え、リスト化
 - 実験者は、死刑についての説明に対する欠陥(CET-for)と、死刑に反対する説明に対する欠陥(CET-against)のそれぞれで、合理的な欠陥の数を数えた
 - Ex.
「情報が明確でない (どのような国を対象にしたのか、どれほどの範囲の影響か、どれぐらい起きているかなどが述べられていない)」
 - ◇ 代替説明の批判的評価能力 : CET-for と CET-against の合計数
 - ◇ 死刑というトピックを選択した理由
 - 先行研究の批判的評価課題において重要なトピックだったため
 - 気候変動のトピックと関連のない方法で批判的評価能力を測定するため
- 手続き
 - 学生は授業の一貫として、自身のコンピュータを使用して、実験を行った
 1. フェイスシート記入
 2. HICCK (プレ)
 3. 妥当性評定 (プレ)

4. 反証的文章を読むグループ (反証グループ: $n = 45$) と、説明的文章を読むグループ (統制グループ: $n = 50$) にランダムに分ける
 5. 文章読後、再び HICCK (ポスト) と妥当性評定 (ポスト) を行う
 6. 批判的評価課題 (CET)
- モデル分析
 - 相違ベースの構造方程式モデリング (VB-SEM) を行う (≡ 共分散構造分析)
 - ◇ 比較的少ないサンプルサイズに適している
 - 指標
 - ◇ GoF (goodness-of-fit; 全体的な適合度)
 - .36 以上のとき、強い説明力を持つ
 - ◇ APC (average path coefficient) と ARS (average R^2 or ARS)
 - モデルの妥当性について、0 からの有意差を示す
 - ◇ AVIF (average variance inflation factor for model parameters) と AFVIF (average full collinearity VIF)
 - 高いほど、潜在的に大きな共線性を示す
 - 変数の必要性を示すために、通常、3.3 以下でないといけない

3. Results

- Table 1 : 評価、妥当性、知識(HICCK)得点の平均と標準偏差
- Table 2 : Model A と Model B の妥当性得点
 - Table 1 における全体の妥当性得点は、Table 2 における Model A から Model B の得点を引いたもの
- Table 3 : 本研究の変数の相関係数

3.1. リサーチクエスション①

- 反証的文章、もしくは説明的文章を読んだとき、批判的評価能力、妥当性についての判断、気候変動知識がどのように関連するのか
 - 反証グループと統制グループのモデルをそれぞれ検討するために、VB-SEM を実施
 - 全体の適合度は両方とも良かった
 - ◇ 反証モデル : GoF = .64
(APC = .459, $p < .001$; ARS = .413, $p < .001$; AVIF = 1.19; AFVIF = 1.87)
 - ◇ 統制モデル : GoF = .54
(APC = .326, $p < .001$; ARS = .306, $p < .001$; AVIF = 1.14; AFVIF = 1.43)
 - 両モデルとも、ポストの HICCK に対して有意に分散を示した (反証モデル: $R^2 = .51$, $p < .001$; 統制モデル: $R^2 = .33$, $p < .001$; Table 3)
 - ◇ このように、両モデルとも、全てのパラメータにおいて十分な効果量を示す
- 両モデルの標準化されたパス図 (Fig. 3)
 - 上図 : 反証モデル

- 下図：統制モデル
- モデル間を比較して興味深い点①
 - 反証モデルにおいて、CET と妥当性 (ポスト) の間に有意な関連が示されたこと ($\beta = .35, p < .01$)
 - 統制モデルにおいて、CET と妥当性 (ポスト) の間が有意傾向であったこと ($\beta = .12, p = .10$)
 - ◇ この結果は、批判的評価能力と文章を読んだ後の妥当性の関連に、反証的文章の影響が予測されることを示唆する
- モデル間を比較して興味深い点②
 - 反証モデルにおいて、ポストの HICCK に対する妥当性判断の標準化係数 ($\beta = .60, p < .01$) は、事前知識 (HICCK pre; $\beta = .19, p = .03$) よりも大きい
 - ◇ 背景知識よりも、知識の妥当性の影響が非常に強い
 - 統制モデルにおいて、ポストの HICCK に対する事前知識 (HICCK pre) の標準化係数 ($\beta = .44, p < .01$) は、妥当性判断 ($\beta = .23, p = .01$) よりも大きい
 - ◇ 妥当性の判断と比較して、相対的に背景知識の影響は強い
- 他のモデルの可能性を考慮し、反証グループにおいて、2つの代替モデルを検討
 - 批判的評価能力がより優れた知識を引き起こし、高い妥当性に繋がるモデル
 - ◇ 全体的な分散は有意だった ($R^2 = .41, p < .001$) が、仮定したモデルより小さく、全体的な適合度と他のモデル指標は、さらに低かった
 - 妥当性が、より優れた批判的評価能力を引き起こし、知識に繋がるモデル
 - ◇ 全体的な分散は有意だった ($R^2 = .33, p < .001$) が、仮定したモデルより小さく、全体的な適合度と他のモデル指標も、仮定したモデルより低かった
- 以上の結果は、仮定したモデルを支持した
 - 反証的文章を読んだ参加者は、批判的評価能力を活性化し、科学的説明に対してより妥当であると判断し、知識テストにおいてより高い得点を示した
 - 説明的文章を読んだ参加者は、潜在的に、批判的評価能力を活性化させず、むしろ、先行知識と初めの妥当性の判断に依存した

3.2. リサーチクエスチョン②

- 反証的文章は妥当性と知識の変化に影響するのか
 - 多変量分散分析 (MANOVA) を実施
 - ◇ 独立変数：文章グループ (反証／統制；参加者間)、時間 (プレ／ポスト；参加者内)
 - ◇ 従属変数：妥当性得点、知識得点 (HICCK)
 - その結果、妥当性と知識を合わせた文章グループと時間の交互作用が有意 ($F(2, 92) = 3.30, p = .041$)

- 妥当性と知識のそれぞれで、一変量分散分析を実施 (Fig.4, Fig. 5)
 - 時間とグループ間の交互作用は、どちらの従属変数においても有意
 - ◇ 妥当性 : $F(1, 93) = 4.24, p = .042$
 - ◇ 知識 : $F(1, 93) = 4.93, p = .029$
 - 単純主効果の分析の結果、それぞれの従属変数で、時間(プレ)における反証/統制グループ間に有意差はなかった (p -values > .68)
 - ◇ しかし、どちらの従属変数にでも、時間(ポスト)において、反証グループは統制グループより有意に高かった
 - 妥当性 : $F(1, 93) = 5.20, p = .025$
 - 知識 : $F(1, 93) = 4.65, p = .033$
 - さらに、統制グループにおける妥当性と知識は、どちらも、文章を読む前後で有意差がなかった (p -values > .84)
 - ◇ しかし、反証グループにおける妥当性と知識は、どちらも、文章を読む前と比較して、読んだ後に有意に増加した (p -values < .01)

4. Discussion

- 仮説 1a,1b は主に支持された
 - 批判的評価能力が、反証的文章を読んだ後の妥当性と知識に有意に関連した
 - 批判的評価能力は、説明的文章を読んだ後の妥当性と知識には関連しなかった
- 先行知識は、両方のモデルにおいて、読後の知識と有意に関連を示した
 - 驚くべきことは、先行知識の影響が、反証的文章によって減少したこと
 - ◇ 反論を含んだ文章を提示することで、参加者が科学的な内容を批判的に評価し、代替説明の妥当性を判断する助けになるかもしれない
- 仮説 2a,2b も支持された
 - 反証的文章を読むことで、批判的評価能力が活性化し、妥当性と知識の両方の変化に関連した
 - ◇ 反証的文章によって評価能力が活性化されるとき、代替説明の妥当性についての判断は知識変化に重要な役割を果たす可能性がある
- 本研究は、妥当性が知識の再構築の重要な構成要素であることを支持する実験的研究の新しい枠組みを示した
 - 妥当性の判断は長期記憶から引き出された情報と結論の両方に影響した
 - 反証的文章は批判的評価能力を活性化し、代替説明についての妥当性を再評価し、知識の再構築を促進した
 - ◇ 評価、妥当性、知識の関係におけるより多くの研究は、学生が科学的事象を深く理解する際の構成概念に対する潜在的な重要性を示すために必要
- 代替説明の妥当性判断と評価は、学生の科学的推論能力の発達を促進するスキル

- 妥当性の再評価を促進する方法として、反証的文章の可能性を主張
 - ◇ 反証的文章は、批判的な材料に関係した学生の先天的な能力を活性化するかもしれない
 - ◇ 教育において、教育的スキルを教える足掛かりとして反証的文章を使うことが可能かもしれない
 - 教育的ツールとしての反証的文章の有用性の検討が必要

- 本研究の限界
 1. 課題時間が短かった
 - ◇ 反証グループにおいて、批判的評価能力と妥当性の再評価の間の潜在的に強固な関連性だけでなく、妥当性と知識の両方の有意な変化を示した
 - 教示において反証的文章を繰り返し使用したことが、効果を持続させたのかもしれない
 2. 一つの科学的トピックを反映した一種類の文章を使用した
 - ◇ 複数の文章や他のトピックを使用したモデルの検討が必要
 3. 批判的評価課題は、批判的評価能力の測定方法の一つ
 - ◇ 潜在的に本研究の一般化可能性を制限
 - ◇ 批判的評価課題は、反証的文章がどのように批判的評価を活性化させるかについて、直接的な原因が推論できない
 4. 現在の気候変動に対する 2 つの代替説明の妥当性のみを測定した
 - ◇ 本研究で使用した説明は、科学的な概念と誤解の主な組み合わせに関連するけれど、妥当性は低いが、他の説明の可能性もある

- 複雑な科学的トピックについて学生の知識を深めることは、しばしば、多くの努力と意図的なプロセスを必要とする (Sinatra & Taasoobshirazi, 2011)
 - その中には、批判的な評価と妥当性の再評価も含むかもしれない
- 本研究は、そのような評価と妥当性の再評価が反証的文章を通して促進されることを示唆し、特に、代替説明をより批判的に見られる個人にその傾向があることを示した
 - 代替説明の妥当性を批判的に再評価することは、科学的に正確な概念の知識を再構築することによって、必要不可欠な構成要素である可能性
 - 反証的文章は、批判的思考と妥当性の判断を教育するツールとして、効果的な手法である可能性

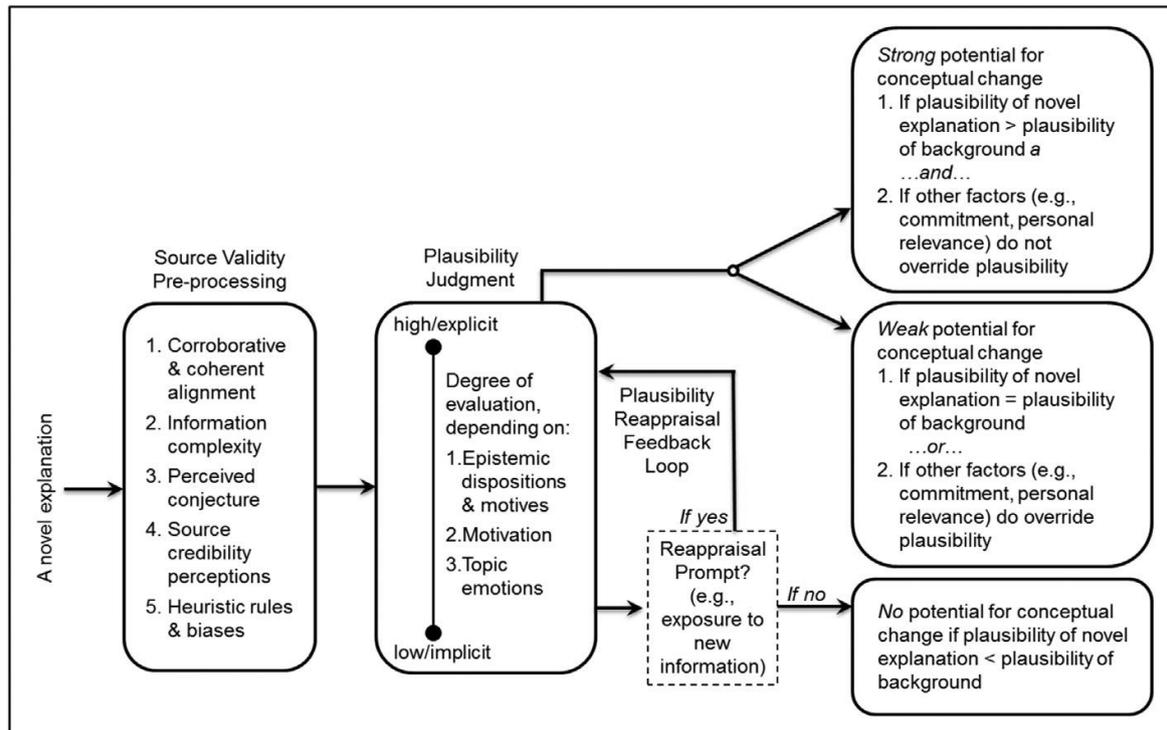


Fig. 1. A model of the role of plausibility judgments in conceptual change (PJCC). From Lombardi, Nussbaum, and Sinatra (2015), *Educational Psychologist*.

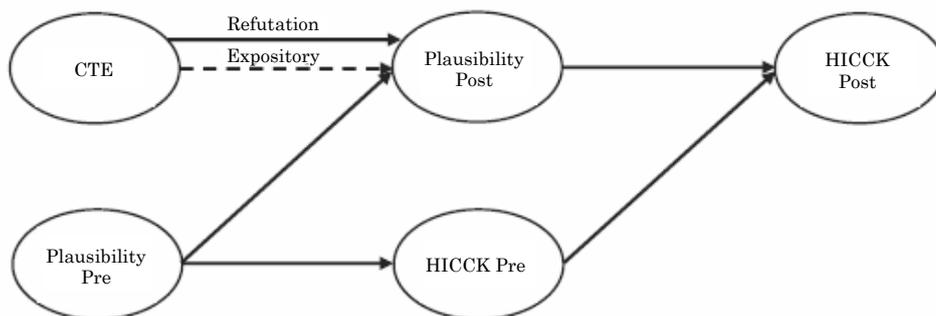


Fig. 2. Hypothetical model of the influence of critical evaluation on plausibility judgments and knowledge. Note. CET = critical evaluation task; HICCK = human-induced climate change knowledge; solid lines indicate potentially significant pathways associated with the refutation effect; and the dashed line indicates a potentially insignificant pathway associated with reading standard expository text. Pre and post for both knowledge and plausibility indicate prior to and after reading, respectively.

Table 1
 Means and standard deviations (in parentheses) for the study variables at pre and post reading ($n_{\text{refutation}} = 45$, $n_{\text{control}} = 50$, $N_{\text{total}} = 95$).

Variable	Group	Pre reading	Post reading
CET	Refutation		2.84 (1.66)
	Control		2.68 (2.22)
	Overall		2.75 (1.97)
Plausibility	Refutation	.78 (3.12)	2.27 (3.86)
	Control	.50 (3.57)	.40 (4.09)
	Overall	.63 (3.35)	1.28 (4.07)
HICCK	Refutation	97.6 (9.72)	102 (12.6)
	Control	97.2 (8.06)	96.9 (11.2)
	Overall	97.4 (8.84)	99.4 (12.1)

Note. The possible score ranges were: (a) critical evaluation task CET = 0 to unlimited (the actual maximum for all participants was 8); (b) plausibility = -9 to +9; and (c) knowledge of human-induced climate change (HICCK) = 27 to 135.

Table 2
 Means and standard deviations (in parentheses) for plausibility scores of the scientific and popular skeptics' alternative models ($n_{\text{refutation}} = 45$, $n_{\text{control}} = 50$, $N_{\text{total}} = 95$).

Model	Group	Pre reading	Post reading
Scientific (human-induced climate change)	Refutation	7.05 (2.06)	7.49 (2.50)
	Control	6.86 (2.29)	6.74 (2.55)
Popular skeptics' (increasing solar energy)	Refutation	6.27 (2.24)	5.22 (2.55)
	Control	6.36 (2.49)	6.34 (2.50)

Note. The possible score ranges were from 1 to 10.

Table 3
 Intercorrelations for the study variables.

Variable	1	2	3	4	5
Refutation text group (n = 45)					
1. HICCK-pre	—				
2. Plausibility-pre	.567**	—			
3. HICCK-post	.460**	.131	—		
4. Plausibility-post	.466**	.402**	.677**	—	
5. CET	-.067	.047	.113	.126	—
Control text group (n = 50)					
1. HICCK-pre	—				
2. Plausibility-pre	.297*	—			
3. HICCK-post	.514**	.274	—		
4. Plausibility-post	.315*	.580**	.381**	—	
5. CET	.144	.104	.210	.178	—

Note. HICCK = human-induced climate change knowledge; CET = critical evaluation task.

* $p < .05$.

** $p < .01$.

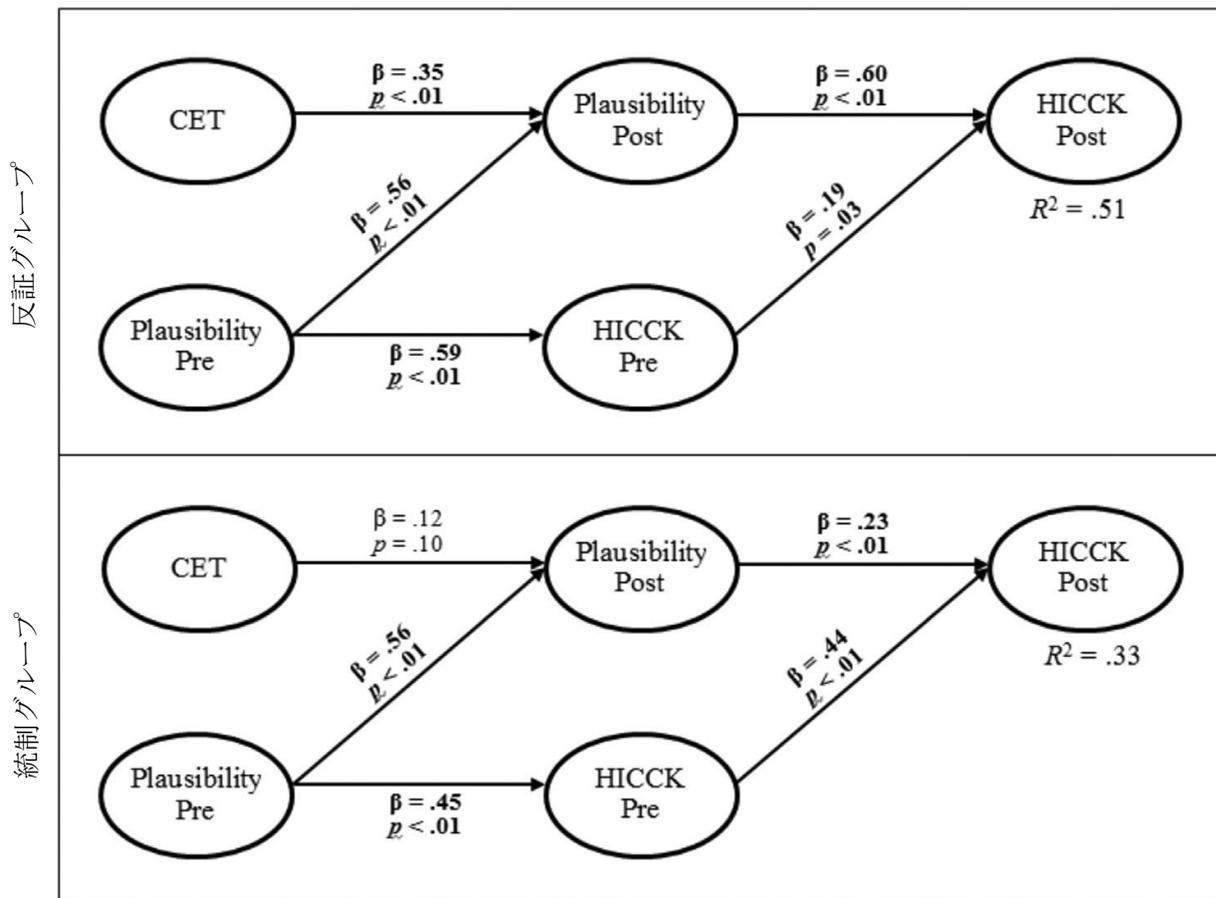


Fig. 3. Models of the influence of critical evaluation on plausibility judgments and knowledge. In the figure, the refutation text model is shown on the top and the control text model is shown on the bottom. Statistically significant standardized path values are shown in bold. Note. CET = critical evaluation task; HICCK = human-induced climate change knowledge.

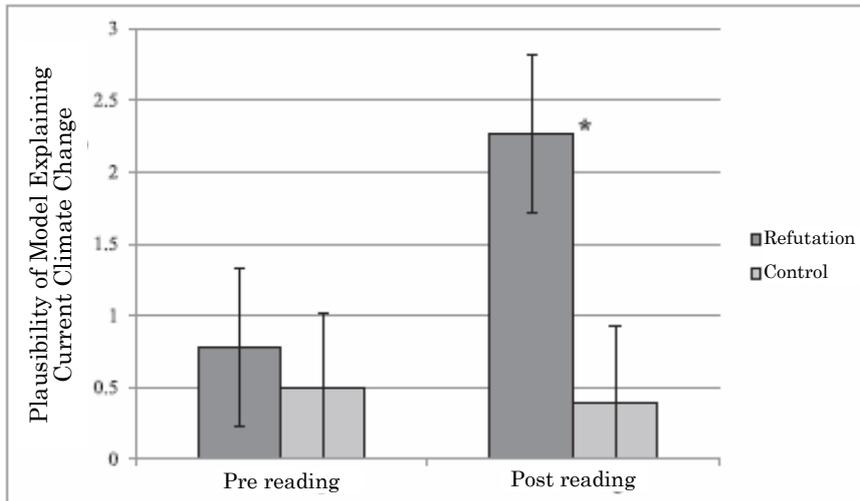


Fig. 4. Model plausibility scores for the refutation and control text groups, with bars showing standard errors. Model plausibility is calculated as participants' plausibility ratings of the scientific model of human-induced climate change minus plausibility ratings of a popular skeptics' model of sun-induced climate change. Therefore, positive scores indicate a greater plausibility toward the scientific model. *Post reading refutation text group plausibility scores were significantly greater than pre reading scores for both groups, as well as post reading scores for the control text group, with all p -values $\leq .025$. Furthermore, there were no significant differences between other scores.

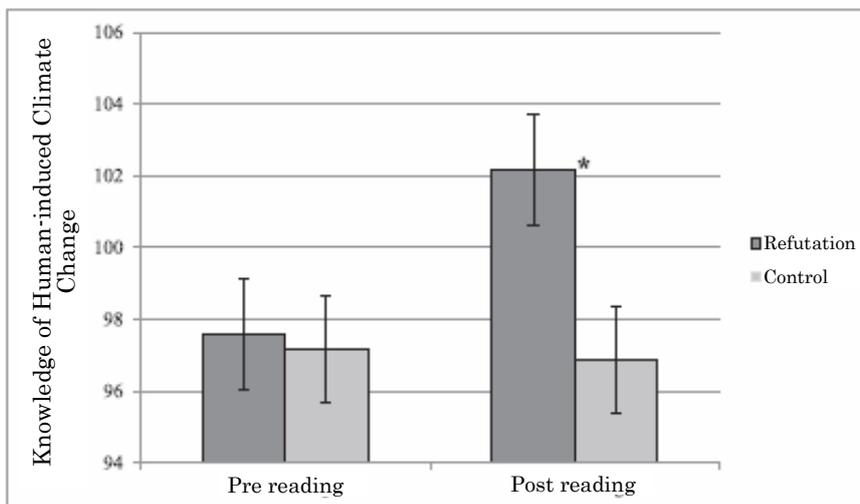


Fig. 5. Human-induced climate change knowledge scores for the refutation and text groups, with bars showing standard errors. *Post reading refutation text group knowledge scores were significantly greater than pre reading scores for both groups, as well as post reading scores for the control text group, with all p -values $\leq .033$. Furthermore, there were no significant differences between other scores.