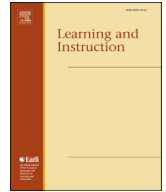




コンテンツリストは ScienceDirect で入手可能

学習と指導

ジャーナルホームページ: www.elsevier.com/locate/learninstruc

問題と例のペアがタスクのパフォーマンスに及ぼす影響 およびさまざまな種類の認知負荷



ジミー・レピンク a、*、フレッド・パース b、c、タマラ・ファン・ゴッグ b、シース P.M.ファン・デル・フルーテン、ジェロエン J.G.ヴァン・メリエンボア

a マーストリヒト大学教育開発研究学部、オランダ b エラスムス・ロッテルダム大学心理学研究所、オランダ c ウーロンゴン大学学際教育研究所、オーストラリア

記事 情報

抽象的な

記事履歴: 2013 年 5 月
30 日受領 修正版受領
2013 年 12 月 2 日
2013 年 12 月 3 日承認

キーワード: 認知負荷
例と例のペア
問題と問題のペア
問題と問題のペア

2 つの研究で、最近開発された心理測定機器が、内因性、外因性、および関連する認知負荷を区別できるかどうかを調査しました。研究 I では、言語学習 (n = 108) と統計講義 (n = 174) について同様の 3 因子ソリューションが明らかになり、統計試験のスコアは、講義中の内因性および外因性の認知負荷を表すと想定される因子と負の相関を示しました。研究 II では、ベイズの定理の応用例 (n = 18) または例問題 (n = 18) の条件で学習した大学 1 年生は、応用を問題例 (n = 18) または問題問題 (n = 20) の条件で学習した同級生よりも、事後テストの成績が優れており、前述の心理測定機器をわずかに修正したバージョンは、研究者が内因性および外因性の認知負荷を区別するのに役立つ可能性があります。この研究結果は、本質的な認知負荷に対処するために使われる実際の作業記憶リソースを指すものとして、関連する認知負荷の最近の再概念化を裏付けるものです。 © 2013 Elsevier Ltd. 無断転載禁止。

1. はじめに

認知負荷理論 (Sweller, 2010 年, Sweller, Ayres, Kalyuga, 2011 年, Sweller, Van Merriënboer, Paas, 1998 年, Van Merriënboer, Sweller, 2005 年, 2010 年) の中心的な教義は、人間の認知構造、特にワーキングメモリの限界を指導の設計時に考慮する必要があるというものです。ワーキングメモリの容量は、単に情報を保持する場合、7 プラスマイナス 2 の情報要素 (またはチャンク) に制限されており (Miller, 1956 年)、情報を処理する場合、さらに少なくなります (約 4) (Cowan, 2001 年)。したがって、ワーキングメモリ負荷 (または認知負荷) は、一定時間内に同時に処理する必要がある情報要素の数によって決まります (Barrouillet, Bernardin, Portrat, Vergauwe, Camos, 2007 年)。もともと、認知負荷理論は、認知負荷の 2 つの原因を区別していました。

内在的認知負荷と外在的認知負荷 (Sweller, 2010 年, Sweller 他, 2011 年, 1998 年)。内在的認知負荷は、学習する情報の内在的性質、より具体的には、学習タスクまたは学習教材を構成する相互作用する情報要素の数によって決まります (Sweller, 1994 年, Sweller 他, 2011 年)。タスクまたは教材に関する事前知識がほとんどない初心者、タスクまたは教材を学習するために、それらの相互作用する要素を処理 (つまり、選択、整理、統合) する必要があります。学習が進むにつれて (つまり、専門知識が増すにつれて)、情報要素は長期記憶に保存される認知スキーマに組み込まれ (またはチャンク化され)、作業記憶で 1 つの要素として処理できるようになります。したがって、学習課題や学習教材によって課される内在的認知負荷は、上級者よりも初心者の方がはるかに高くなります。無関係な認知負荷は、学習者が認知スキーマの構築に直接貢献しない認知プロセス (たとえば、空間的または時間的に分離されているが相互に参照する情報源を精神的に統合する必要がある) に従事することを要求する、最適ではない指導方法から生じ、学習目標には不要で無関係です (Sweller & Chandler, 1994; Sweller, Chandler, Tierney, Cooper, 1990)。このようなプロセスは、内在的認知負荷が高い場合や、

* 連絡先著者。マーストリヒト大学教育開発研究部、P.O. Box 616, 6200 MD マーストリヒト、オランダ。電話: +31 433885709。電子メールアドレス: jimmie.leppink@maastrichtuniversity.nl (J. Leppink)。

内在的認知負荷が低い状況では、最適とは言えない学習につながる。つまり、そのような状況では、学習を妨げることなく外在的認知負荷を管理できるとしても、外在的認知負荷を学習に直接関連する認知負荷（すなわち、適切な認知負荷、Sweller 他、1998 年）に置き換えると、より良い学習成果が得られるはずである。適切な認知負荷の概念は、後に認知負荷フレームワークに追加されました（Sweller 他、1998 年）。このタイプの負荷は、長期記憶またはコンテキストからの関連情報を新しい情報要素に関連付けることから生じ（Sweller、2010 年、Sweller 他、2011 年）、内在的認知負荷に対処するために割り当てられたワーキングメモリリソースに関係します（Kalyuga、2011 年、Sweller、2010 年）。実際、「関連する認知負荷」という用語は、認知負荷理論の従来の概念化で使用されてきた（Sweller et al., 1998）一方、「関連するリソース」（つまり、内在的認知負荷に対処するために割り当てられたワーキングメモリリソース）という用語は、理論の最近のバージョンで使用されており、内在的認知負荷に関連しています（Kalyuga、2011; Sweller、2010; Sweller et al., 2011）。認知負荷理論によれば、学習者の事前知識や習熟度に一致する教材を選択することで、指導設計において内在的認知負荷を最適化する一方で、無関係な認知負荷は最小限に抑え、学習者は、（認知負荷理論の古い概念化では）適切な認知負荷を喚起するプロセス、または（理論の新しい概念化では）適切なリソース（例えば、実践、詳細化、または自己説明における変動性）の使用に取り組むように促され、認知スキーマの構築に直接貢献する（Sweller et al., 1998; Van Merriënboer & Sweller、2005、2010）。両方の用語を互換的に使用することで理論の2つの異なる概念化を参照することによる混乱を避けるため、本稿の残りの部分では、Kalyuga (2011)、Sweller (2010)、および Sweller et al. (2011) が示唆したように、関連するリソースの使用を指す用語として「関連する認知負荷」を使用します。

1.1. 指導指導と認知負荷

指導上の特徴が内在的または外在的認知負荷にどの程度寄与するかは、個々の学習者によって異なります。たとえば、情報によって内在的認知負荷が高い初心者学習者は、自律的な問題解決よりも、実例（つまり、完全に解決された問題の解決策、Cooper & Sweller、1987、Paas、1992、Paas & Van Merriënboer、1994a、Sweller & Cooper、1985、Van Gog、Paas、& Van Merriënboer、2006）などの外在的認知負荷を軽減する指導形式や、部分的に解決された解決策（つまり、完了問題、Paas、1992、Van Merriënboer、1990）を完了することからよりよく学習する可能性があります。問題解決は初心者学習者にとって大きな無関係な認知負荷を課します。なぜなら、そのタイプの問題を解決する方法に関する事前の知識がないため、彼らは弱い問題解決戦略に頼らざるを得ないからです。解決策（の一部）は実例と補完問題で解決されるため、弱い問題解決戦略の使用によって課される無関係な認知負荷は防止され、学習者は作業記憶リソースの多くを内在的認知負荷（つまり、関連リソース）に対処するために割り当てることができます。一方、知識が豊富な学習者は、そのタイプの問題を解決する方法に関する知識をすでに獲得しており、それが問題解決の指針となるため、自律的な問題解決から最大限の利益を得られます。初心者学習者にとって有益な指導形式は、その効果を失い、知識が豊富な学習者にとってはマイナスの結果をもたらすことさえあります（専門知識の逆転効果、Kalyuga、Ayres、Chandler、Sweller、2003 年、Kalyuga、Chandler、Tuovinen、Sweller、2001 年、Leppink、

Broers、Imbos、Van der Vleuten、Berger、2012a、2012b、2013b）。実例で提示される情報は、知識が豊富な学習者にとっては冗長であり、そのような学習者は指導なしで問題を解くことができ、冗長な情報の処理は余分な認知負荷につながります（つまり、冗長性効果、Chandler および Sweller、1991）。

1.2. 主観的評価尺度による認知負荷の測定

Paas (1992) の 9 段階の一次元精神努力評価尺度などの主観的評価尺度は、学習者が経験する全体的な認知負荷を測定するために頻繁に使用されてきました（レビュー: Paas、Tuovinen、Tabbers、& Van Gerven、2003、Van Gog & Paas、2008）。精神努力は Paas 氏によって「タスクによって課される要求に対応するために実際に割り当てられる認知能力であり、したがって実際の認知負荷を反映していると考えられる」と定義されています（Paas、Tuovinen ら、2003、p. 64。Paas & Van Merriënboer、1994b も参照）。作業負荷と認知負荷が、文脈を超えてどの程度同じ概念を指すのかは完全には明らかではありませんが、多次元 NASA-TLX (Hart & Staveland、1998) は、5 つの 7 段階評価スケールで経験作業負荷を主観的に評価する別のツールの例です。各ポイントの高、中、低の推定値の増分により、スケールには 21 段階の段階があります（Hilbert & Renkl、2009; Zumbach & Mohraz、2008）。主観的または客観的な手法で全体的な経験認知負荷を測定することは、特に学習成果の測定に関連して有益ですが（Van Gog & Paas、2008）、指導の設計に情報を提供するという点では、異なるタイプの認知負荷を個別に測定するよりも具体的ではありません。そのため、いくつかの研究では、3 種類の認知負荷を個別に測定する手段の開発が試みられてきました（Ayres、2006 年、Gierniak、Scheiter、Gerjets、2009 年、De Leeuw、Mayer、2008 年、Eysink ら、2009 年、Galy、Cariou、Mélan、2012 年）。これらの研究の欠点は、1 つ以上の種類の認知負荷が 1 つの項目で表されていることです。認知負荷の異なる種類ごとに複数の指標を使用すると、より正確な測定が可能になり、研究者は各尺度に 1 つの指標を使用するよりも、認知負荷の種類をより明確に区別できるようになります。さらに、非常に具体的な指導上の特徴や認知プロセスを参照して、無関係な認知負荷または関連する認知負荷を測定する場合、概念上の問題が生じる可能性があります。専門知識の逆転効果により、特定の指導上の特徴が、ある学習者にとっては関連する認知負荷と関連し、別の学習者にとっては無関係な認知負荷と関連している可能性があることが示されるためです（Kalyuga ら、2001、2003）。

1.3. 認知負荷の3つのタイプを区別するための新しい測定機器

最近、異なるタイプの認知負荷を測定するための質問の作成に別のアプローチを採用した心理測定機器が開発されました（Leppink、Paas、Van der Vleuten、Van Gog、Van Merriënboer、2013）。これにより、異なるタイプの認知負荷を少なくともある程度区別できないという問題が解決される可能性があります。Sweller (2010) と Kalyuga (2011) が最近示唆したように、関連する認知負荷が内在的認知負荷に対処するために割り当てられた作業記憶リソースに関係する場合、関連する認知負荷と内在的認知負荷を区別することは困難である可能性があります。この新しい心理測定機器（Leppink、Paas 他 2013）は堅牢な 3 因子構造を明らかにしましたが、いくつかの理由から、これらの 3 つの因子が実際に 3 種類の認知負荷を表しているかどうかはまだ明らかではありません。

まず、ランダム化実験における関連認知負荷とその後のタスクパフォーマンスとの相関は予想よりも低く、統計的に有意ではありませんでした。次に、Leppink、Paas ら (2013) が発表した一連の研究はすべて、統計教育という 1 つのコンテキストに焦点を合わせていました。3 つの要因が実際に 3 種類の認知負荷、またはこれらの種類の認知負荷に関連する安定した構成を表す場合、これらの要因が他のコンテキストでも表面化すると予想されます。3 番目に、Leppink、Paas ら (2013) が適用した実験操作では、3 つの要因のいずれにも期待された違いは実際には生じませんでした。

1.4. 現在の研究

私たちは、前述の心理測定機器 (Leppink、Paas et al., 2013) を使用して 2 つの研究を実施し、(1) この機器が内在的、外在的、および関連する認知負荷を区別するのに役立つかどうか、(2) その機器から得られた要因がタスクパフォーマンスの予測因子として使用できるかどうか、(3) これらの要因が指導の設計によってどのように影響を受けるか (研究 II: 問題-問題、問題例、例-問題、および例-例のペアに焦点を当てる) を調査しました。

2. 研究 I: 言語と統計の探索的分析

研究 I では、統計学の講義で最初に開発されテストされたツールを言語の授業に適応させました。表 1 は、研究 I で使用された 2 つのバージョンの質問票を示しています。言語の授業バージョンは、語彙と統計概念、および文法と統計式に関連付けて作成されたことに注意してください。コミュニケーションをとるためには、統計知識領域で重要な概念と定義に関する十分な知識が不可欠です。これは、言語を話すには語彙に関する十分な知識が不可欠であるのと同じです。さらに、文法と式の両方に知識と適用ルールが必要です。したがって、統計学習用と言語学習用の 2 つのバージョンのツールが、3 つの要素のそれぞれについて同様の 3 因子パターンと同等の内部一貫性値を明らかにすると予想されました。さらに、統計学の講義は、試験で完了するコースの一部でした。試験の成績は、適切な認知負荷を表すとされる要因と正の相関があり、内在的および外在的認知負荷を表すとされる要因と負の相関があると予想しました。したがって、研究 I では 3 つの仮説が検証されました。言語と統計の領域における認知負荷測定用の新しい 2 つのバージョンは、3 つの要因のそれぞれについて、類似した 3 因子パターンと同等の内部一貫性値を生み出す (H1)、統計の領域における試験の成績は、内在的および外在的認知負荷を表すとされる要因と負の相関がある (H2)、試験の成績は、適切な認知負荷を表すとされる要因と正の相関がある (H3)。

2.1. 方法

2.1.1. 参加者と材料 表 1 に示す言語クラス版の調査票は、言語コースを選択した学生が参加した合計 14 の言語クラス (n = 108) で実施されました (言語とヨーロッパ共通参照枠 [CEFR] レベルに関する情報については表 2 を参照、欧州評議会、2011 年)

表 1 研究 I で使用した「認知負荷」質問票の 2 つのバージョン。項目 1 から 3 は内在的認知負荷を捕捉するもの、項目 4 から 6 は外在的認知負荷を捕捉するもの、項目 7 から 10 は関連認知負荷を捕捉するものと想定されていました。

統計 次の 10 の質問はすべて、終了したばかりの講義に関するものです。時間をかけて各質問を注意深く読み、提示された 0 から 10 のスケールで各質問に回答してください ('0' はまったく当てはまらないことを示し、'10' は完全に当てはまることを示します): 0 1 2 3 4 5 6 7 8 9 10 [1] 講義で扱われたトピックは非常に複雑でした。[2] 講義で扱われた数式は、非常に複雑だと感じました。[3] 講義で扱われた概念と定義は、非常に複雑だと感じました。[4] 講義中の指示と説明は非常に不明瞭でした。[5] 講義中の指示と説明は、不明瞭な言葉でいっぱいでした。[6] 講義中の指示と説明は、学習という点では、非常に効果がありませんでした。[7] 講義によって、扱われたトピックの理解が非常に深まりました。[8] 講義によって、扱われた数式に対する理解が非常に深まりました。[9] 講義によって概念と定義に関する私の知識が本当に深まりました。[10] 講義によって主題に関する私の知識と理解が本当に深まりました。

言語 次の 10 の質問はすべて、終了したばかりのレッスンに関するものです。時間をかけて各質問を注意深く読み、提示された 0 から 10 のスケールで各質問に回答してください ('0' はまったく当てはまらないことを示し、'10' は完全に当てはまることを示します): 0 1 2 3 4 5 6 7 8 9 10 [1] レッスンで扱われたトピックは非常に複雑でした。[2] レッスンで扱われた文法構造は、非常に複雑だと感じました。[3] レッスンで扱われた語彙は、非常に複雑だと感じました。[4] レッスン中の指示と説明は非常に不明瞭でした。[5] レッスン中の指示と説明は、不明瞭な言葉でいっぱいでした。[6] レッスン中の指示と説明は、学習の点では非常に効果ありませんでした。[7] レッスンによって、扱われたトピックの理解が本当に深まりました。[8] このレッスンで、学んだ文法構造の理解が本当に深まりました。[9] このレッスンで、語彙の知識が本当に深まりました。[10] このレッスンで、言語に関する知識と理解が本当に深まりました。

学習カリキュラムの選択科目として、または別個のコースとして。統計版の調査票は、学士課程 1 年生の心理学科の学生 (n = 174) を対象とした推論統計コースの講義で実施された。トピック

表 2 研究 I における言語クラスごとの言語、目標 CEFR レベル、生徒数 (合計 108 人)。

グループ 言語 目標 CEFR レベル 生徒数

1 フランス語 A2 7 2 オランダ語 A2 9 3 オランダ語 B2 6 4a イタリア語 A1 7 5 スペイン語 A2 8 6 ポルトガル語 A1 6 7 ドイツ語 A1 8 8a イタリア語 B1 7 9 中国語 A1 9 10 ロシア語 A1 7 11 スペイン語 A1 8 12 フランス語 B1 10 13 オランダ語 A1 6 14 フランス語 B2 10

同じ先生です。

講義で扱われた標本平均の標本分布（および標準誤差などの関連概念）は、統計コースの中核をなしており、帰無仮説有意性検定、母平均に関する仮説検定の検定統計量 z 、 t 、 F 、および母集団比率に関する仮説検定の検定統計量 $c2$ の概念も取り上げられています。コースのすべてのトピックと、コース終了時の試験のすべての質問は、この特定の講義の内容に直接関連していました。標本分布、標準誤差、および標本理論の関連概念を理解していない学生は、帰無仮説有意性検定と検定統計量の論理を理解できず (Ben-Zvi & Garfield, 2004; Leppink, Broers, Imbos, Van der Vleuten, & Berger, 2011; Leppink, Broers, Imbos, Van der Vleuten, & Berger, 2013a)、これらの概念の理解を必要とする試験に不合格になる可能性が高くなります。

2.1.2. 手順 言語クラスは小グループで行われ、パフォーマンスデータは収集できませんでした。アンケートデータは、異なる CEFR レベルで教えられた異なる言語クラス (表 2 を参照) および異なる教師によって収集されました。このようなアプローチでは、言語、教師、および/または言語レベル間でツールの結果を比較することはできませんが (これはこの研究の目的ではありません)、ツールの心理測定特性を判定できるデータを収集することはできます。学生は 2 時間の授業の最後にアンケートに回答しました。統計の講義は 2 時間続き、学生は講義の最後にアンケートに回答するように指示されました。174 人の学生全員が講義の最後に紙のアンケートに回答し、すぐに提出しました。これらの学生のほとんどは 5 週間後にコース試験を完了しました ($n = 151$)。Leppink, Paas らは、2010 年 11 月 1 日にコースの試験を完了した学生の 1 人あたりのアンケートの回答者数を 100 人に増やしました。(2013) は、質問項目の順序を変えても因子負荷量や内部一貫性の値に大きな影響がないことを実証したため、私たちは 10 項目を表 1 に示す順序で提示することにしました。項目の順序は両方のアンケートで同じであることに注意してください。

2.1.3. データ分析 言語コースの学生は学習グループ内に入れ子にされていました。このようなマルチレベル設計は、グループ内および学生間の相関構造を誘発し、学習グループ間で変数 (ここでは質問票の因子) 間の関係が変化する可能性があります。ただし、おそらくクラス数とクラス内の学生数が限られているため、このようなクラス内係数は小さいか無視できる程度で、統計的に有意ではありませんでした。さらに、言語クラスのサンプル総数 ($n = 108$) は確認的因子分析には小さすぎましたが、経験則によると、参加者の数が質問票の項目数の 10 倍 (つまり 10) であれば、グループは探索的因子分析に十分な大きさでした。そのため、次のように進めました。主因子分析を実施し、両方の質問票バージョンについて因子ごとの内部一貫性値を計算して、言語と統計の領域における認知負荷測定用の新しいツールの 2 つのバージョンが、3 つの因子のそれぞれについて同様の 3 因子パターンと同等の内部一貫性値を生成するかどうかをテストしました (H1)。さらに、統計領域での試験成績が、内在的および外在的認知負荷を表すと想定される因子と負の相関関係にあるかどうか (H2)、および関連する認知負荷を表すと想定される因子と正の相関関係にあるかどうかをテストするために (H3)、統計講義データとコース試験から得られた 3 つの因子間の相関関係を計算しました。これらの相関関係は、コース試験を完了した 151 人の学生 (講義の回答者 174 人のうち 86.8%) のデータに基づいています。

2.2. 結果

記述統計では、異常な回答パターンや極端なケースは見られませんでした。どちらの学習状況でも、項目への回答は 0 から 10 までの全範囲またはほぼ全範囲をカバーし、ほとんどの項目は絶対歪度と尖度の値が [$\diamond 1.5$; 1.5] または [$\diamond 1$; 1] の範囲内にありました。統計コースの試験スコアについても同じことが言えます。これは、それぞれ 1 つの正解選択肢がある合計 17 の多肢選択式質問に対する正解数の合計スコアでした (理論的な範囲は 0 から 17 です)。平均試験スコアは 10.30 で、標準偏差は 2.98 でした。両方のアンケートのデータは主因子分析に適していました。値には十分な相互相関があり、Kaisere MeyereOlkin (KMO) のサンプリング適正度は統計では .814、言語では .748 と良好な値であり、Bartlett の球面性検定は統計的に有意で、統計では $c2(45) = 1223.962$, $p < .001$ 、言語では $c2(45) = 616.651$, $p < .001$ でした。3 因子解が予想され、因子間の相関が予想されたこと (Leppink, Paas ら、2013 に基づく) を考慮して、因子の相互相関を考慮するために斜交 (Oblimin) 回転が実行されました。両方の質問票で 3 因子解が見つかり、統計の総分散の 80.8 パーセントと言語の総分散の 75.6 パーセントを説明しました。両方のコンテキストで、最初の因子に負荷された関連認知負荷を表す項目、2 番目の因子に負荷された内在的認知負荷を表す項目、3 番目の因子に負荷された外在的認知負荷を表す項目がそれぞれありました。表 3 は、2 つの質問票のそれぞれの因子負荷と内部一貫性 (因子ごとの Cronbach のアルファ値) を示し、表 4 は斜交回転の根拠として 3 つの因子間の相関関係を示しています。2 つのバージョンは、同様の因子負荷と内部一貫性値を生み出します。因子相関関係は、Leppink, Paas らによって報告されたものと同様です。(2013): 外的認知負荷と関連認知負荷を表すと想定される因子間の相関は負、内的認知負荷と外的認知負荷を表すと想定される因子間の相関は正、内的認知負荷と関連認知負荷を表すと想定される因子間の相関はほぼゼロ。平均スコアは各学生の因子ごとに計算された。表 5 は、アンケートごとの 3 つの因子のそれぞれについて、平均、標準偏差、歪度、尖度を示している。統計分野の試験成績は、内的認知負荷と外的認知負荷を表すと想定される因子と負の相関があるという仮説 (H2) に沿って、試験成績は因子と負の相関があった。

表 3 研究 I における因子ごとの因子負荷量と内部一貫性 (つまり、Cronbach のアルファ) 値。

要素/項目	統計	言語
読み込み中 アルファ 読み込み中 アルファ		
「本質的」	.893	.816
項目 1	.932	.893
項目 2	.782	.734
項目 3	.846	.688
「無関係」	.785	.838
項目 4	.693	.925
項目 5	.911	.881
項目 6	.569	.557
「関連性」	.947	.889
項目 7	.922	.904
項目 8	.858	.635
項目 9	.905	.809
項目 10	.933	.909

表 4 研究 I における 3 つの要因間の相関関係。

因子ペア 統計 言語	
「本質的」と「外的」	.504 .271
「本質的」と「関連」	◆.125 .029
「外的」と「関連」	◆.210 ◆.519

講義中の内的認知負荷 ($r = .210$, $p = .010$) と外的認知負荷 ($r = .320$, $p < .001$) を表すと想定される。試験の成績は、関連する認知負荷を表すと想定される要因と正の相関関係にあるという仮説 (H3) に沿って、試験の成績は、講義中の関連する認知負荷を表すと想定される要因と正の相関関係にあった ($r = .140$, $p = .084$) が、この相関関係は、従来の $\alpha = .05$ レベルでは統計的に有意ではなかった。

2.3. 議論

Leppink, Paas 他 (2013) による因子構造の類似した文脈 (統計教育) と異なる文脈 (言語学習) での複製は、3 つの因子が堅牢であることを示唆しています。ただし、これらの因子が 3 種類の認知負荷を表すかどうかはまだ明らかではありません。試験の成績と、関連する認知負荷を表すとされる因子との相関は正でしたが、その相関は小さく、統計的に有意ではありませんでした。さらに、内的認知負荷と外的認知負荷を表すとされる因子間の正の相関が何を意味するのか疑問視することもできます。Sweller 他 (1998) が示唆するように、内的、外的、関連する認知負荷が実際に独立した、加算的なタイプの認知負荷である場合、これらのタイプの認知負荷間の相関は (ほぼ) ゼロであると予想されます。その観点から見ると、内因的負荷と外因的負荷を表すとされる要因間の正の相関は、回答者がそれらを区別するのが難しい可能性があることを示唆している可能性があります。内因的認知負荷と外因的 (および関連のある) 認知負荷を区別する試みがこれまであまり成功しなかったのは、部分的にはこのためかもしれません。一方、私たちは事前の知識を制御しておらず、心理学の学生は中等教育の軌跡に応じて数学の事前知識が異なる可能性があるため、タスクの内因的認知負荷が高かった学生は外因的認知負荷も高かった可能性があります。さらに、試験の成績と内因的認知負荷と外因的認知負荷を表すとされる 2 つの要因間の負の相関は、理論的予測と実証的発見と一致しています。外因的 (つまり、非効率な) 認知負荷が高いと、学習が妨げられることが予想されます (Sweller および Chandler, 1994 年)。さらに、タスクの内的認知負荷 (つまり、タスクの複雑さ) が高いほど、学習成果が低くなる傾向があること (Ayres, 2006)、および内的認知負荷は事前知識の個人差によって影響を受けること (Sweller, 2010; Sweller et al., 2011, 1998) が実証されています。したがって、内的認知負荷と学習成果の負の相関は、

表 5 研究 I のアンケートにおける 3 つの因子ごとの平均 (M) と標準偏差 (SD)、および歪度と尖度。

因子	統計 M (SD)	歪度	尖度	言語 M (SD)	歪度	尖度
「本質的」	4.63 (2.03)	◆.138	◆.862	4.45 (1.90)	◆.080	◆.092
「無関係」	2.09 (1.55)	1.571	4.076	1.74 (1.58)	.930	◆.116
「関連性」	6.30 (1.73)	◆.710	1.006	7.13 (1.64)	◆.699	.847

認知負荷と試験の成績は、統計的概念 (および/またはこれらの概念の基礎となる数式) の習得が一般的に難しい学生は、より高い内的認知負荷を経験し、試験の成績が悪かったことを反映している可能性があります。内的認知負荷と外的認知負荷を表すはずの 2 つの要因が実際にこれら 2 種類の認知負荷を表す場合、これは、研究者が内的認知負荷と外的認知負荷を区別するのに役立つ心理測定機器を開発したことを意味します。ただし、研究 I では実際には予想していなかった別の可能性があります (Leppink, Paas ら 2013 も同様です)。項目への回答は、(予想される) タスクの複雑さの推定または認識、(予想される) 指導の非効率性の原因の反映、および理解と知識の獲得の反映を反映している可能性があります。これらはこれら 3 つの要因に実際に費やされた努力とは無関係であり、その場合、作業記憶のどの側面からも独立しています。内的および外的認知負荷を捉えると想定される項目は、内的および外的認知負荷に実際に費やされた努力の指標ではなく、必要な内的および外的認知負荷活動の推定を反映している可能性があります。同様の推論が関連認知負荷を捉えると想定される項目にも当てはまる可能性があります。これらの考えられる説明をさらに調査し、経験したさまざまな種類の認知負荷と学習成果とのより直接的な関係を調べるために、研究 II で、タスクの形式と順序を変え、参加者が初心者となるトピック (ペイズの定理) を使用し、表 6 に示す 3 つの項目を表 1 に示す元の 10 項目に追加する実験が行われました。3 つの要因が実際に内的、外的、および関連認知負荷を表している場合、各要因 e に実際に費やされた精神的努力について尋ねる 3 つの項目 e が、問題の要因の内部一貫性に寄与すると予想されます。

3. 研究 II: 問題、例、タスクパフォーマンス、認知負荷

研究 II では、ランダム化実験が行われました。社会科学と健康科学の 1 年生が、問題、問題例、問題例、または問題例の条件で、初心者である難しいトピックであるペイズの定理の応用について学習し、学習後とその後のタスク実行後に、心理測定機器の修正版を使用して認知負荷を測定しました。この研究では、Van Gog, Kester, および Paas (2011) による研究結果を別の領域で再現することもできます。彼らの研究は、1 つの実験内でこれらの 4 つの指導条件を比較した最初の研究であり、興味深い結果が得られました。まず、彼らの研究は、問題例と問題例の条件を比較し、問題例と問題例の条件でテストのパフォーマンスに有意差がなく、学習フェーズに費やされた精神的努力に有意差がないことを実証した最初の研究でした。言い換えれば、実例を学習した後に問題を解くことは、別の例を学習する場合と比較してテストのパフォーマンスを向上させませんでした。Van Gog と Kester (2012) は、即時テストに関するこの発見を再現し、1 週間後の遅延テストでは、exampleexample 条件のパフォーマンスが exampleproblem 条件よりもさらに優れていることを実証しました。次に、Van Gog ら (2011) の研究では、exampleexample と exampleproblem の両方が problemexample と problemproblem よりも優れたテストパフォーマンスをもたらし、さらに、exampleexample と exampleproblem 条件では、学習フェーズでより少ない精神的努力でこのパフォーマンスに到達しました。

表 6 3 つの新しい項目を含む「認知負荷」質問票。項目 4 (内在的認知負荷を捉える項目 1~3 と同様)、項目 8 (無関係な認知負荷を捉える項目 5~7 と同様)、および項目 13 (関連した認知負荷を捉える項目 9~12 と同様) は、研究 II で新しく追加された項目です。

以下の 10 の質問はすべて、終了したばかりのアクティビティに関するものです。時間をかけて各質問を注意深く読み、提示された 0 から 10 のスケールで各質問に回答してください (「0」はまったく当てはまらないことを示し、「10」は完全に当てはまることを示します): 0 1 2 3 4 5 6 7 8 9 10 [1] このアクティビティの内容は非常に複雑でした。[2] このアクティビティで扱われた問題は非常に複雑でした。[3] このアクティビティでは、非常に複雑な用語が使用されました。[4] このアクティビティの複雑さに非常に多くの精神的努力を費やしました。[5] このアクティビティの説明と指示は非常に不明瞭でした。[6] このアクティビティの説明と指示には不明瞭な言葉が満載でした。[7] このアクティビティの説明と指示は、学習の点では非常に効果がありませんでした。[8] 私はこの活動で、不明瞭で効果のない説明と指示に非常に多くの精神的努力を費やしました。[9] この活動により、カバーされた内容に対する私の理解が本当に高まりました。[10] この活動により、カバーされた問題に対する私の理解が本当に高まりました。[11] この活動により、言及された用語に関する私の知識が本当に高まりました。[12] この活動により、カバーされた問題に対処する方法に関する私の知識と理解が本当に高まりました。[13] この活動中、私は自分の知識と理解を高めるために非常に多くの精神的努力を費やしました。

問題例と問題問題という条件では、生徒は問題例と問題問題という条件で同じ結果を得ました。この結果は、実例効果に関する多数の研究と一致しています (レビューについては、Sweller 他、1998 年、Van Gog と Rummel、2010 年を参照)。3 番目に、問題例と問題問題という条件では、テストの成績や学習段階での精神的努力に違いはありませんでした。最初は、問題例という条件の方が問題問題という条件よりも成績が良かったという結果は意外かもしれません。結局のところ、この 2 つの条件の生徒は、条件が異なるだけで、同じ量の指導サポートを受けたのです。しかし、この結果は Reisslein、Atkinson、Seeling、Reisslein (2006) の研究結果とも一致しており、例を最初に学習することで、後で問題を解決するときに使用できる認知スキーマ (つまり、適切な認知負荷アクティビティ) を構築できるため、条件が重要であることを示唆しています (問題のある例の条件と比較して、問題の固有の認知負荷が低下します)。最初に問題を解決する場合、無関係な認知負荷が高く、その問題を解決することで得られる学習がほとんどない可能性があります。ただし、後者では、問題のある例の条件で、問題のある例の条件よりもパフォーマンスが向上しなかった理由をまだ説明していません。繰り返しになりますが、Van Gog ら (2011) の研究は、これら 2 つの条件を比較した最初の研究であり、この結果は予想外でした。Van Gog らによって提供された考えられる説明は次のとおりです。問題は、学生が問題解決の経験に非常に苛立ち、その後の例をあまりよく勉強しないということです。これは、例の問題条件と問題の例条件の間の適切な認知負荷評価の差に反映されるはずです。したがって、これらの調査結果が別の領域で再現できるかどうかを確認するだけでなく、新しい機器で経験した認知負荷を測定して、この調査結果のパターンをより適切に説明できるかどうかを判断することも興味深いでしょう。これらの調査結果に沿って、例の問題または例の例条件でベイズの定理の適用を学習した初心者参加者は、ベイズの定理の適用に関するその後の事後テストで、他の参加者よりも良い成績を収めると予想されました。

問題例または問題問題条件 (H4a) でこのアプリケーションを学習した参加者と比較して、2 番目の形式は事後テストのパフォーマンスに有意な影響を与えなかった (H4b)。認知負荷に関しては、各要因に実際に費やされた精神的努力を尋ねる認知負荷ツールに追加された 3 つの項目が、内在的認知負荷 (H5a)、無関係な認知負荷 (H5b)、および関連認知負荷 (H5c) を表すと想定される要因の内部一貫性に寄与すると仮定した。さらに、問題例と問題問題条件の学生は、問題例と問題問題条件の学生よりも、事後テストでの内在的認知負荷 (H6) が低く、学習段階と事後テストでの無関係な認知負荷が低く (H7a と H7b)、学習段階と事後テストでの関連認知負荷が高い (H8a と H8b) と報告すると予想した。

3.1. 方法

3.1.1. 参加者と実験設計 社会科学と健康科学の学部1年生 84名が、問題、問題例、問題例、例例の条件にランダムに割り当てられました。最初のタスクは、自律的な問題解決または実例の学習のいずれかを伴い、2番目のタスクも同様です。潜在的なコンテキスト効果を考慮するために、異なるコンテキストで2つの問題を設計しました。2つのコンテキストは、カウンターバランスの取れた順序で提示されました。つまり、参加者の約半数が1つのコンテキストで最初のタスクを完了し、残りの参加者は別のコンテキストで最初のタスクを完了し、その逆も同様です。9名の学生は、授業スケジュールの変更により直前に参加をキャンセルし、もう1名の学生は指示に従わなかったため、実験から除外されました。この結果、次の状況が発生しました: problemeproblem (n ¼ 20)、problemeeexample (n ¼ 18)、exampleeproblem (n ¼ 18)、exampleeeexample (n ¼ 18)。

3.1.2. 材料と手順 3.1.2.1. 認知負荷測定器。研究 II では、3 つの要因が実際に何を表しているかをよりよく理解するために、いくつかの項目をわずかに改訂し、認知負荷測定器に 3 つの項目を追加しました (表 6 を参照)。

3.1.2.2. 学習教材とテスト教材。2つの問題 (つまり、文脈) は、ベイズの定理の同じ応用に焦点を当てていました。

PðAjbB 1/4PðAð ◆PðBjAð◆=PðBð

このトピックが選ばれた理由は 2 つあります。まず、参加者全員がカリキュラム内の後続の統計コースでこのトピックを学習することになります。実験に参加した直後に、この定理の応用に関する 1 時間の講義が行われ、参加者が学習意欲を掻き立てられる現実的な教育環境が提供されます。次に、統計は科学研究だけでなく、他の職業や日常生活でも重要なツールです。統計は主に経験的現象の数学的モデリングに関するものであり、不確実性の科学であり、条件付き確率がすべての統計の中核を形成します。

3.1.3. 手順 手順は以下のとおりです。実験は30分間続きました。参加者は最初のタスク (自律的な問題解決、または最初のコンテキストまたは2番目のコンテキストでの例) を紙に5分で実行し、最初のタスクを提出した後、2番目のタスク (自律的な問題解決、または最初のコンテキストまたは2番目のコンテキストでの例) に5分かかりました。どちらのタスクも、

条件付き確率を導く計算は行われず、自由回答形式の質問とみなすこともできます（複数選択の選択肢は提供されていません）。タスクを自律的に実行する必要がある参加者は、この条件付き確率を自分で提供する必要がありますが、実例を学習した参加者は、問題の解決策として正しい計算と条件付き確率を認識できます。最初の 10 分後、学生は調整された「認知負荷」アンケート（13 項目）に初めて回答し、この 10 分間のトレーニングフェーズで経験した認知負荷を評価しました。5 分間の休憩後、学生は 2 つの新しいコンテキスト（つまり、コンテキストごとに 3 つの質問）に基づいて作成された 6 つの自由回答形式の質問で構成される事後テストを受けましたが、ベイズの定理の適用はまったく同じでした。質問は、学習フェーズで自律的な問題解決を必要とした問題とまったく同じストーリーラインに沿っており、つまり、参加者は自分で正しい条件付き確率を計算して提供しなければなりませんでした。コンテキストと質問は、誤ったアルゴリズム（例えば、結合確率 $P(A, B)$ を計算したり、正しい $P(A|B)$ の代わりに条件付き確率 $P(B|A)$ を提供したり）を使用して正しい条件付き確率を提供することが不可能であり、正しい答えを推測することが事実上不可能であるように作成されました。このようにして、正しい応答は、参加者がベイズの定理を正しく適用できることを反映していると確信できます。したがって、正しい応答の合計（つまり、0e6 のスケールの整数）が非常に高い Cronbach のアルファ（ $\alpha \approx .950$ ）を生成し、すべての項目の p 値（つまり、正しい応答の割合）が .70e.78 の範囲にあり、修正された項目と全体の相関が .695 以上であることは驚くことではありません。これにより、学生がベイズの定理を適用できる範囲を非常に正確に測定でき、指導条件が参加者の事後テストの成績に与える影響に対する最適な統計的検出力が得られました。すべての学生は、事後テストに与えられた 15 分以内に 6 つの質問すべてを完了することができました。事後テストを完了した後、すべての学生は調整された「認知負荷」質問票（13 項目）を再度完了し、事後テストで経験した認知負荷を評価しました。

3.1.4. データ分析 指導条件（すなわち、4 つの条件）が事後テストのパフォーマンスに与える影響を調べるために、2 元配置の被験者間（BS）分散分析（ANOVA）を実施しました。これにより、次の 3 つの特定の対比をテストできます。（1）第 1 タスクの自律的な問題解決と実例（すなわち、第 1 タスクの種類の主効果）、（2）第 2 タスクの自律的な問題解決と実例（すなわち、第 2 タスクの種類の主効果）、および（3）1 つの特定の条件の追加効果（すなわち、第 1 タスクと第 2 タスクの種類の相互作用効果）。BS ANOVA を 3 元配置に拡張して、モデルにバランスのとれたトレーニング コンテキストを考慮に入れても、事後テストのパフォーマンスの説明には役立ちませんでした。バランスのとれたトレーニング環境は、事後テストの成績にわずかしが影響せず、統計的に有意ではなかった。 $F(1, 72) = .891, p = .348, \eta^2 = .012$ （.01、.06、.14 の値は、それぞれ小さい、中程度、大きい効果を示す。Field, 2013）。したがって、2 元配置 BS ANOVA で十分だった。最初の形式（H4a）の主効果は統計的に有意だが、2 番目の形式（H4b）の主効果は統計的に有意ではなく、相互作用効果も統計的に有意ではないと予想した。「認知負荷」質問票（H5a、H5b、H5c）の 3 つの追加項目の付加価値を調べるために、3 つの各因子について、因子ごと、測定機会ごと（つまり、トレーニングと事後テスト）のクロンバックのアルファ、および修正された項目対全体の相関と、特定の項目を削除した場合のクロンバックのアルファ値を計算しました。次に、平均スコアを計算しました。

測定機会ごとの因子ごとに、事後テストのパフォーマンスと測定機会ごとの（平均）因子スコアとの相関関係を調査しました。実例から始める学生は、事後テスト（H6）での内在的認知負荷が低く、学習段階（H7a）と事後テスト（H7b）の両方で外在的認知負荷が低く、学習段階（H8a）と事後テスト（H8b）で関連認知負荷が高いという仮説を検証するために、2元配置分散分析と同じBS因子（第1および第2タスク）を使用し、2回の測定機会（トレーニングと事後テスト）からの各因子のスコアを反復測定（被験者内、WS）として扱い、分割プロット分散分析を実行しました。後者は、BS と WS の相互作用をテストすることを可能にします（これは、一部の条件が事後テストでは固有の認知負荷が異なるが、学習段階では異なる場合に存在します）。また、学習段階と事後テストのパフォーマンスについてこれらの主な効果を別々にテストするよりも、第1形式と第2形式の主効果をテストするためのパワーが高くなります。最後に、事後テストのパフォーマンスの場合と同様に、分析にカウンターバランスされたトレーニング コンテキストを含めることは、どの因子スコアの説明にも寄与していませんでした。すべての h^2 値は .005 から .014 の範囲にあり（事後テストのパフォーマンスの場合の h^2 値 .012 に相当）、 p 値は .309 から .542 の範囲でした。

3.2. 結果

事後テストのパフォーマンスは、やや左に偏っており（歪度 $\frac{1}{4}$ ◆1.195）、平均スコア（M）は 4.49、標準偏差（SD）は 2.34 でした。問題→問題という条件のパフォーマンスが最も悪く（M $\frac{1}{4}$ 3.50、SD $\frac{1}{4}$ 2.78）、次いで問題→例という条件（M $\frac{1}{4}$ 4.11、SD $\frac{1}{4}$ 2.52）、例→問題という条件（M $\frac{1}{4}$ 5.06、SD $\frac{1}{4}$ 1.83）となり、例→例という条件のパフォーマンスが最も高くなりました（M $\frac{1}{4}$ 5.39、SD $\frac{1}{4}$ 1.65）。予想どおり、第 1 タスクと第 2 タスク間の相互作用効果は統計的に有意ではありませんでした（ $F(1, 70) \frac{1}{4}$.070、 $p \frac{1}{4}$.793、 $h^2 \frac{1}{4}$.001）。これは、第 1 タスクと第 2 タスク（存在する場合）の効果は、加法的と見なすことができることを意味します。第 2 タスクで実例を学習した参加者（つまり、問題例または例例）は、第 2 タスクで問題を自律的に解決した参加者よりも平均 .472 ポイント（範囲: 0e6）高く評価しましたが、この差は統計的に有意ではなく、 $F(1, 70) \frac{1}{4}$.805、 $p \frac{1}{4}$.373、 $h^2 \frac{1}{4}$.011 であり、効果サイズは、せいぜい小さな効果について話していることを示しています。ただし、第 1 タスクの効果は統計的に有意であり、 $F(1, 70) \frac{1}{4}$ 7.245、 $p \frac{1}{4}$.009、 $h^2 \frac{1}{4}$.094 であり、効果サイズは、中程度からやや大きい効果を示しています。最初のタスクで実例を学んだ参加者は、最初のタスクで問題を自主的に解いた参加者よりも、事後テストではるかに良い成績（0e6スケールで1.417ポイント）を収めました。

表7は、条件ごとの測定機会(トレーニングと事後テスト)ごとの内在的、外在的、および関連した認知負荷を表すと考えられる因子の(平均)スコアの平均(M)と標準偏差(SD)に加えて、歪度と尖度を示しています。

表 8 は、各測定機会ごとに、修正された項目対全体の相関関係と、特定の項目を削除した後（残りの項目は保持）の Cronbach のアルファ値を示し、表 9 は、事後テストのパフォーマンスとの相関関係を示しています。

表8は、3つの要因が訓練と事後テストで同等のクロンバックのアルファ値を示したことを示しているが、訓練の余分な認知負荷を表すとされる要因のクロンバックのアルファ値はわずかに低く、これは範囲制限効果を反映している可能性がある。表9は、事後テストのパフォーマンスと、訓練の余分な認知負荷を表すとされる3つの要因との相関関係を示している。

表7 指導形式ごとの測定機会(トレーニングと事後テスト)ごとの内在的、外在的、および関連した認知負荷を表すと考えられる因子の(平均)スコアの平均(M)と標準偏差(SD)、および歪度と尖度(研究II)。

因子 トレーニング M (SD) 歪度 尖度 事後テスト M (SD) 歪度 尖度									
問題例「本質的」	2.03 (1.44)	.646	◆.575	2.26 (1.83)	.786	.151	「無関係」	1.09 (1.36)	
1.284 .550	1.49 (1.76)	1.095	◆.234	「関連」	2.03 (1.84)	.487	◆1.267	1.94 (1.74)	.599
◆.291	問題例「本質的」	1.78 (1.61)	.749	◆.537	3.10 (1.97)	.608	1.105	「無関係」	1.14
(1.30)	1.180	1.025	1.43 (1.48)	1.264	◆.413	「関連がある」	1.74 (1.47)	.361	◆.994
2.42	(1.98)	.678	◆.942	例題「本質的」	2.60 (1.85)	.584	◆.623	2.65 (1.47)	.498
◆.913	「無関係」	1.54 (1.49)	.772	◆.782	1.86 (1.47)	.503	◆.048	「関連がある」	4.12 (2.43)
◆.125	◆1.305	4.56 (2.11)	◆.247	◆.533	例「本質的」	2.32 (1.62)	.046	◆.778	2.79 (1.89)
.392	◆.767	「無関係」	2.01 (1.72)	.870	.007	2.07 (1.74)	.642	◆.959	「関連」
.526	◆.566	3.00 (1.73)	.178	.200					

内因性、外因性、および適切な認知負荷を表す因子は、2回の測定で類似しており、(特に最初の2つの因子については)ゼロに近い。最後に、表10から12は、それぞれ内因性、外因性、および適切な認知負荷を表すと想定された3つの因子の分割プロットANOVAの結果を示している。表10から12のいずれも、形式と時間の間、または最初の形式と2番目の形式の間に統計的に有意な相互作用効果を示していない。表12の最初の形式のみが統計的に有意であり、 h^2 値が.170であることは大きな効果を示している。

3.3. 議論

Van Gog ら (2011) の研究と一致して、例と例の問題のペアは、問題と問題の例のペアよりもポストテストのパフォーマンスが優れていることがわかりました (H4a)。また、2番目の形式はポストテストのパフォーマンスに大きな影響を与えませんでした (H4b)。表8に示されている結果は、認知負荷ツールに追加された3つの項目のうち2つ(各項目に実際に費やされた精神的努力について尋ねる)が、

表8 クロンバックのアルファ値、補正された項目全体相関、および特定の項目を削除した後のクロンバックのアルファ値(残りの項目は保持)、各測定機会ごとの要因、および事後テストのパフォーマンスとの相関(研究II)。

要因/項目 トレーニング 事後テスト			
修正済み 項目合計 相関	クロンバックのアルファ 要素 項目 アウト 要素 項目 アウト	クロンバックのアルファ	
「本質的」	.853	.872	項目 1 .849 .740 .833
.790	項目 2 .819 .755 .876 .770	項目 3 .493	
.886	.465 .921	項目 4 .662 .826 .764 .821	
「無関係」	.632	.787	項目 5 .451 .537 .725
.661	項目 6 .525 .562 .622 .755	項目 7 .363	
.631	.431 .837	項目 8 .469 .531 .734 .663	
「関連」	.933	.931	項目 9 .921 .897 .894
.899	項目 10 .892 .903 .891 .900	項目 11	
.826	.917 .834 .913	項目 12 .802 .922 .848	
.909	項目 13 .680 .942 .637 .948		

因子eの修正は、内在的認知負荷(H5a)と外在的認知負荷(H5b)を表すと想定される因子の内部一貫性に寄与する。これら2つの因子に関して、修正された項目対総相関(および問題の項目を除いたCronbachの α 値)は、(内在的または外在的特徴において)実際に費やされた精神的努力について尋ねる追加項目eが内部一貫性に寄与することを示す。これは、内在的認知負荷を捉えると想定される項目が1つの同じ潜在的構成を測定し、外在的認知負荷を捉えると想定される項目が1つの同じ潜在的構成を測定することを示していると思われる。問題の2つの構成は、内在的認知負荷と外在的認知負荷、またはこれら2種類の認知負荷に関連する因子である可能性がある。適切な認知負荷を表すと想定される3番目の因子については、結果はそれほど説得力がないように見える。表12に示されている結果は、例と例の問題の条件の学生は、問題の例と問題の条件の学生よりも、学習段階(H8a)と事後テスト(H8b)でより高い関連認知負荷を報告するという仮説を裏付けています(h^2 値が.170で、大きな効果を示しています)。ただし、表8に示されている結果は、関連認知負荷活動に実際に費やされた精神的努力について尋ねる項目が、関連認知負荷を表すと想定される因子の内部一貫性に寄与するという仮説を裏付けていません(H5c)。追加項目の修正された項目対全体の相関は依然としてかなり高いですが、この値は他の項目の値とは区別できます。これは、他の4つの項目が関連認知負荷活動に実際に費やされた努力を直接的に捉えていないことを反映している可能性があります。これは、研究IおよびLeppink, Paas ら (2013) の実験で、関連する認知負荷活動を表すとされる因子と事後テストの成績との間にわずかな相関関係が見られ、研究IIではそれほど高くないこと(表9を参照)を説明するかもしれません。これらの相関関係は再び正ですが、その大きさは、3番目の因子と関連する認知負荷の関係がせいぜい限られていることを示しているようです。

表9 測定機会ごとの3つの要因と事後テストのパフォーマンスとの相関関係(研究II)。

因子 トレーニング r (p 値) 事後テスト r (p 値)
「本質的」 .068 (.567) ◆.017 (.883) 「無関係」 .049 (.677) .042 (.722) 「関連性」 .141 (.230) .113 (.336)

表 10 内在的認知負荷を表すと想定される因子の分割法分散分析の結果 (研究 II)。

効果 F(1, 70) p値 h2値
時間 7.121 .009 .092 最初のタスク .731 .395 .010 2 番目のタスク .101 .751 .001 最初のタスクと 2 番目のタスクの差 1.204 .607 .004 最初のタスクの時間 1.728 .193 .024 2 番目のタスクの時間 3.715 .058 .050 最初のタスクと 2 番目のタスクの差 .737 .393 .010

異なる条件で、内在的および外在的認知負荷を表すとされる要因の平均スコアに差がないという結果は予想外でした。これは、例と例の問題の条件の学生が事後テストで内在的認知負荷が低いと報告する (H6) という仮説や、例と例の問題の条件の学生が、問題と問題の例の条件の学生よりも学習段階 (H7a) と事後テスト (H7b) で外在的認知負荷が低いと報告するという仮説と一致しません。おそらく、学習段階が短すぎて内在的または外在的認知負荷に有意な影響を与えなかったか、または、実例の有益な効果はむしろ適切な認知負荷によるものなのかもしれません。しかし、この観点からすると、表 12 に示されている結果が、例と例の問題の条件の学生は、問題の例と問題の問題の条件の学生よりも、学習段階と事後テストでより高い関連認知負荷を報告するという仮説を裏付けていることは、おそらくさらに興味深いことです。これは、学生がこの「知識と理解」要因を評価し、条件間で異なる方法で判断できることを示している可能性があります。そうだとすれば、関連負荷そのものでなくても、それは依然として潜在的に重要な構成概念です。

4. 一般的な議論

これらの調査結果を総合すると、内因性認知負荷と外因性認知負荷は心理測定機器を使用して区別できる、または少なくとも内因性認知負荷と外因性認知負荷を表すとされる要因は内因性認知負荷と外因性認知負荷に関連しているという仮説をある程度裏付けているように思われます。2 つの要因は研究全体で一貫して現れ、実際に費やされた努力について尋ねる新しく追加された項目は 2 つの要因それぞれの信頼性に寄与し、これらの要因に負荷をかける質問は内因性認知負荷と外因性認知負荷の理論的概念に関連している可能性があります。これらの調査結果は、2 つの要因がそれぞれ内因性認知負荷と外因性認知負荷を表すか密接に関連しているという仮説を裏付けているように思われますが、この仮説の妥当性を調べるにはさらなる実験が必要です。研究 II で使用された、内因的および外因的認知負荷を表すとされる 2 つの要因を、研究 II と同様に既知の効果を再現する一連の新しい実験で使用することをお勧めします。これらの要因が、

表 11 外的認知負荷を表すと想定された因子の分割プロット ANOVA の結果 (研究 II)。

効果 F(1, 70) p値 h2値
時間 2.635 .109 .036 最初のタスク 3.323 .073 .045 2 番目のタスク .276 .601 .004 最初のタスクと 2 番目のタスクの差 .285 .595 .004 最初のタスクの時間 .232 .631 .003 2 番目のタスクの時間 .321 .573 .005 最初のタスクと 2 番目のタスクの差 .056 .814 .001

表 12 関連する認知負荷を表すと想定された因子の分割プロット ANOVA の結果 (研究 II)。

効果 F(1, 70) p値 h2値
時間 3.251 .076 .044 最初のタスク 14.337 <.001 .170 2 番目のタスク 2.782 .100 .038 最初のタスクと 2 番目のタスクの差 3.632 .061 .049 最初のタスクの時間 .029 .865 <.001 2 番目のタスクの時間 .722 .398 .010 最初のタスクと 2 番目のタスクの差 1.643 .204 .023

これらが極めて高い内因性および/または外因性認知負荷を課することはほぼ確実であり、これらの要因がそのような状況でのタスクパフォーマンスとどのように相関するかを確認する必要があります。特に、トレーニング中に外因性認知負荷に有意差が見られなかったという事実は驚くべきことであり、さらに調査する必要があります。一方では、これは、私たちの機器が実際には外因性認知負荷を測定していないことを示唆している可能性があります。他方では、実例効果の理論的説明は、外因性認知負荷効果よりも、関連のある認知負荷にある可能性があります。また、学習フェーズが短すぎて内因性または外因性認知負荷に有意な影響を与えなかった可能性を調査するには、研究 II の複製ではより長い学習フェーズを使用する必要があります。内因性認知負荷と外因性認知負荷を表すとされる 2 つの要因が相関しているという事実は、学生が 2 種類の認知負荷を区別するのに多少の困難を抱えていることを反映しているようです。問題の一部は、特定の質問の文言効果にある可能性があります。不明瞭な指示は、必ずしも追加的で無関係な処理から生じるわけではなく、少なくともある程度は事前の知識の欠如から生じる可能性があり、指示は多くの認知活動を伴うため、学習者にとって複雑になる可能性があります。そのうちのいくつかは無関係である可能性があります。今後の実験で、このような文言効果をテストする価値があります。この結果は、最近提案された、本質的な認知負荷に対処するために割り当てられた実際の作業記憶リソースを指す、関連する認知負荷の再概念化と一致しているようです (Kalyuga, 2011 年, Sweller, 2010 年, Sweller 他, 2011 年)。少なくとも、この結果は、そのような再概念化に反対するものではありません。心理測定機器の 3 番目の要因が関連する認知負荷を表している、または密接に関連しているという仮定に対する支持は限られています。投入された努力に関する追加項目は、この要因の信頼性に寄与しなかった。これは、もともと適切な認知負荷を捉えたと想定されていた項目が、実際には割り当てられた特定のタイプの作業記憶リソースを指しているという仮定を裏付けるものではない。研究 II の調査結果は、学生が「知識と理解」要因を評価し、条件によって異なる判断を下す可能性があることを示している可能性がある。そうだとすれば、それ自体は適切な負荷ではないとしても、それは依然として潜在的に重要な構成概念である。この要因とパフォーマンスの相関が一貫して小さいことの 1 つの説明は、学習者は学習であると認識しているものに多くの努力を費やしたり、ほとんど費やさなかったりするが、その努力の一部は無駄になる可能性があるということである可能性がある。この文脈における 2 番目の説明は、追加のフィードバックがなければ、学生は努力の関連性を限られた範囲でしか振り返ることができず、おそらくより上級の学生やドメインの専門家だけが実際にそうすることに成功するということである可能性がある。後者は、さまざまなドメインの初心者と専門家を含む新しい研究で調査する価値がある。新しいトピックの自主学習を、自律的な問題解決ではなく、実例から始めると、タスクのパフォーマンスに良い効果があるという発見は、実例効果に関する以前の研究 (Cooper & Sweller, 1987 年, Paas, 1992 年, Paas & Van Merriënboer, 1994a, Sweller & Cooper, 1985 年, Van Gog et al., 2006 年) や、問題例と実例の関連性に関する以前の研究と一致しています。

例題ペア (Reisslein et al., 2006; Van Gog et al., 2011)。この発見は、指導ガイダンスに関する最近の議論 (Clark, Kirschner, Sweller, 2012など) に照らして特に興味深いものです。これは、ガイダンスがどれだけ提供されるかだけでなく、いつ提供されるかも重要であることを示しているからです。研究 II を別のトピックで、初心者と上級学習者の両方で再現し、(1) 学習者の習熟度が上がるにつれて、問題と問題および/または問題と例のペアが、例と問題と例のペアよりも効果的になるかどうか (これは、専門知識の逆転効果に関する調査結果と一致します。Kalyuga ら、2001、2003 年、Leppink, Broers ら、2012a、2012b、2013b)、(2) 内在的および外在的認知負荷を表すとされる要因が、異なる指導条件における異なるレベルの専門知識に対してどのように動作するかを調べるのは興味深いでしょう。訴訟事例と実証研究は、共通点がほとんどないように見えることがよくありますが、少なくとも 1 つ共通点があります。ストーリーの妥当性は、証拠の連鎖に関するものです (Kane, 2006 年)。訴訟事件の証拠を扱う場合でも、実証研究を扱う場合でも、私たちは仮定を立てますが、絶対的な証拠は存在しません。刑事事件の容疑者がたった 1 つの証拠に基づいて有罪判決を受けるべきではないのと同様に、測定機器の有効性は 1 つまたは 2 つの研究で確立されるものではありません。それは証拠の連鎖を探る旅であり、その証拠の連鎖を得るには、機器の一部の要素を修正または調整する必要がある場合があります。適切な認知負荷活動を、内在的認知負荷に対処するために割り当てられた作業記憶リソースとして再定義するという最近の提案は、認知負荷理論の透明性と簡潔さの点で非常に魅力的に思われます。Leppink, Paas ら (2013) の調査結果と現在の 2 つの調査結果を合わせると、この動きには実証的な裏付けがあるようです。内在的認知と外在的認知の測定のさらなる発展は、現在定義されている認知負荷理論 (Kalyuga, 2011年、Sweller, 2010年、Swellerら、2011年) と、本論文でいくつかの提案が提示されている実証研究との間の継続的な対話によって推進されるべきである。

参考文献

- Ayres, P. (2006). 主観的尺度を使用して問題内の内在的負荷の変動を検出する。Learning and Instruction, 16, 389e400. Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., Camos, V. (2007). 作業記憶における時間と認知負荷。Journal of Experimental Psychology: Learning, Memory, and Cognition, 33, 570e585. <http://dx.doi.org/10.1037/02787393.33.3.570>. Ben-Zvi, D., Garfield, J. B. (2004). 統計的リテラシー、推論、思考力の育成の課題。ドルドレヒト: Kluwer Academic Publishers. Chandler, P., Sweller, J. (1991). 認知負荷理論と指導の形式。認知と指導, 8, 293e332. <http://dx.doi.org/10.1207/s1532690xci0804.2>. Cierniak, G., Scheiter, K., Gerjets, P. (2009). 分割注意効果の説明: 無関係な認知負荷の減少は、関連する認知負荷の増加を伴うか? Computers in Human Behavior, 25, 315e324. <http://dx.doi.org/10.1016/j.chb.2008.12.020>. Clark, R. E., Kirschner, P. A., Sweller, J. (2012). 生徒を学習の道に導く: 完全ガイド付き指導の事例。American Educator, 36, 6e11. <http://www.aft.org/pdfs/americaneducator/spring2012/Clark.pdf>. Cooper, G. A., Sweller, J. (1987). スキーマ獲得とルール自動化が数学の問題解決の転移に与える影響。Journal of Educational Psychology, 79, 347e362. <http://dx.doi.org/10.1037/0022-0663.79.4.347>. Cowan, N. (2001). 短期記憶における魔法の数字 4: 精神的記憶容量の再考。Behavioral and Brain Sciences, 24, 152e153. De Leeuw, K. E., Mayer, R. E. (2008). 認知負荷の 3 つの尺度の比較: 内在的、外在的、および関連負荷の分離可能な尺度の証拠。Journal of Educational Psychology, 100, 223e234. <http://dx.doi.org/10.1037/0022-0663.100.1.223>. Eysink, T. H. S., De Jong, T., Berthold, K., Kollöffel, B., Opfermann, M., および Wouters, P. (2009). マルチメディア学習アレンジメントにおける学習者のパフォーマンス: 指導アプローチ間の分析。American Educational Research Journal, 46, 1107e1149. <http://dx.doi.org/10.3102/0002831209340235>. Field, A. (2013). IBM SPSS 統計を使用した統計の発見 (第 4 版)。ロンドン: Sage.
- Galy, E., Cariou, M., Mélan, C. (2012). 精神的負荷要因と認知負荷タイプの関係は何か? International Journal of Psychophysiology, 83, 269e275. <http://dx.doi.org/10.1016/j.ijpsycho.2011.09.023>. Hart, S. G., Staveland, L. E. (1988). NASA-TLX (タスク負荷指数) の開発: 実証的および理論的研究の結果。P. A. Hancock, N. Meshtaki (編)、人間の精神的負荷 (pp. 139e183)。アムステルダム、オランダ: North-Holland. Hilbert, T. S., Renkl, A. (2009). コンピューターベースのコンセプトマッピングツールの使い方を学ぶ: 自己説明的な例が役立つ。Computers in Human Behavior, 25, 267e274. <http://dx.doi.org/10.1016/j.chb.2008.12.006>. Kalyuga, S. (2011). 認知負荷理論: 実際に必要な負荷の種類はいくつですか? Educational Psychology Review, 23, 1e19. <http://dx.doi.org/10.1007/s10648-010-9150-7>. Kalyuga, S., Ayres, P., Chandler, P., および Sweller, J. (2003). 専門知識の逆転効果。Educational Psychologist, 38, 23e31. <http://dx.doi.org/10.1207/S15326985EP3801.4>. Kalyuga, S., Chandler, P., Tuovinen, J., Sweller, J. (2001). 問題解決が事例の学習よりも優れている場合。Journal of Educational Psychology, 93, 579e588. <http://dx.doi.org/10.1037/0022-0663.93.3.579>. Kane, M. (2006). 検証。R. L. Brennan (編)、Educational measurement (4th ed.) (pp. 17e64)。Westport: ACE/Praeger. Leppink, J., Broers, N. J., Imbos, Tj., Van der Vleuten, C. P. M., Berger, M. P. F. (2011). 命題操作法 (MPM) における課題および学生関連の要因の調査。統計教育ジャーナル, 19. オンライン <http://www.amstat.org/publications/jse/v19n1/leppink.pdf>. Leppink, J., Broers, N. J., Imbos, Tj., Van der Vleuten, C. P. M., および Berger, M. P. F. (2012a). 事前の知識が統計の概念的理解に対する指導効果を緩和する。教育研究および評価, 18, 37e51. <http://dx.doi.org/10.1080/13803611.2011.640873>. Leppink, J., Broers, N. J., Imbos, Tj., Van der Vleuten, C. P. M., および Berger, M. P. F. (2012b). 統計の領域における自己説明: 専門知識の逆転効果。高等教育, 63, 771e785. <http://dx.doi.org/10.1007/s10734-0119476-1>. Leppink, J., Broers, N. J., Imbos, Tj., Van der Vleuten, C. P. M., および Berger, M. P. F. (2013a). 統計知識領域における講義方法としての命題操作の有効性。Instructional Science. <http://dx.doi.org/10.1007/s11251-013-9268-3>. Leppink, J., Broers, N. J., Imbos, Tj., Van der Vleuten, C. P. M., および Berger, M. P. F. (2013b). 統計の問題ベース学習における指導の効果。ジャーナル 実験教育の教授。 <http://dx.doi.org/10.1080/00220973.2013.813365>. Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., Van Merriënboer, J. J. G. (2013年)。さまざまなタイプの認知負荷を測定するための機器の開発。行動研究方法。 <http://dx.doi.org/10.3758/s13428-0130334-1>. ミラー, G. (1956). 魔法の数字 7 プラスマイナス 2: 情報処理能力の限界。The Psychological Review, 63, 81e97. <http://dx.doi.org/10.1037/0033-295X.101.2.343>. Paas, F. (1992). 統計学における問題解決スキルの転移を達成するためのトレーニング戦略: 認知負荷アプローチ。Journal of Educational Psychology, 84, 429e434. <http://dx.doi.org/10.1037/0022-0663.84.4.429>. Paas, F., Tuovinen, J., Tabbers, H., および Van Gerven, P. (2003). 認知負荷理論を前進させる手段としての認知負荷測定。Educational Psychologist, 38, 63e71. <http://dx.doi.org/10.1207/S15326985EP3801.8>. Paas, F., & Van Merriënboer, J. J. G. (1994a). 事例の変動性と幾何学の問題解決スキルの転移: 同族負荷アプローチ。Journal of Educational Psychology, 86, 122e133. <http://dx.doi.org/10.1037/00220663.86.1.122>. Paas, F., & Van Merriënboer, J. J. G. (1994b). 複雑な認知タスクの訓練における認知負荷の指導制御。Educational Psychology Review, 6, 51e71. <http://dx.doi.org/10.1007/BF02213420>. Reisslein, J., Atkinson, R. K., Seeling, P., および Reisslein, M. (2006)。電気回路解析におけるコンピュータベースの環境での専門知識の逆転効果の発見。Learning and Instruction, 16, 92e103. <http://dx.doi.org/10.1016/j.learninstruc.2006.02.008>. Sweller, J. (1994)。認知負荷理論、学習困難、および教育設計。Learning and Instruction, 4, 295e312. [http://dx.doi.org/10.1016/0959-4752\(94\)90003-5](http://dx.doi.org/10.1016/0959-4752(94)90003-5). Sweller, J. (2010). 要素の相互作用と内在的、外在的、および関連的な認知負荷。Educational Psychology Review, 22, 123e138. <http://dx.doi.org/10.1007/s10648-010-9128-5>. Sweller, J., Ayres, P., Kalyuga, S. (2011). 認知負荷理論。ニューヨーク: Springer. Sweller, J., Chandler, P. (1994). なぜ一部の教材は学習が難しいのか。Cognition and Instruction, 12, 185e223. <http://dx.doi.org/10.1207/s1532690xci1203.1>. Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). 技術資料の構造化における要因としての認知負荷。Journal of Experimental Psychology, 119, 176e192. <http://dx.doi.org/10.1037/0096-3445.119.2.176>. Sweller, J., & Cooper, G. A. (1985). 代数学習における問題解決の代替としての例題の使用。Cognition and Instruction, 2, 59e89. <http://dx.doi.org/10.1207/s1532690xci0201.3>. Sweller, J., Van Merriënboer, J. J. G., Paas, F. (1998). 認知アーキテクチャと教育設計。教育心理学レビュー, 10, 251e296. <http://dx.doi.org/10.1023/A:1022193728205>. Van Gog, T., Kester, L. (2012). テスト効果のテスト: 事例から問題解決スキルを習得する。認知科学, 36, 1532e1541. <http://dx.doi.org/10.1111/cogs.12002>.

Van Gog, T., Kester, L., & Paas, F. (2011). 実例、例題、問題例のペアが初心者の学習に与える影響。Contemporary Educational Psychology, 36, 212e218. <http://dx.doi.org/10.1016/j.cedpsych.2010.10.004>. Van Gog, T., & Paas, F. (2008). 教育の効率: 教育研究における元の構成概念の再検討。Educational Psychologist, 43, 16e26. <http://dx.doi.org/10.1080/00461520701756248>. Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). プロセス指向の実例がトラブルシューティングの移行パフォーマンスに与える影響。Learning and Instruction, 16, 154e164. <http://dx.doi.org/10.1016/j.learninstruc.2006.02.003>. Van Gog, T., およびRummel, N. (2010)。例に基づく学習: 認知研究と社会認知研究の観点の統合。Educational Psychology Review, 22, 155e174. <http://dx.doi.org/10.1007/s10648-010-9134-7>。

Van Merriënboer, J. J. G. (1990). 高校におけるプログラミング指導の戦略: プログラム完成とプログラム生成。Journal of Educational Computing Research, 6, 265e285. <http://dx.doi.org/10.2190/4NK5-17L7-TWQV-1EHL>. Van Merriënboer, J. J. G., Sweller, J. (2005). 認知負荷理論と複雑学習: 最近の動向と今後の方向性。Educational Psychology Review, 17, 147e177. <http://dx.doi.org/10.1007/s10648-005-3951-0>. Van Merriënboer, J. J. G., Sweller, J. (2010). 医療専門職教育における認知負荷理論: 設計原則と戦略。Medical Education, 44, 85e 93. <http://dx.doi.org/10.1111/j.1365-2923.2009.03498.x>. Zumbach, J., Mohraz, M. (2008). ハイパーメディア読解における認知負荷: テキストの種類と直線性の影響。Computers in Human Behavior, 24, 875e887. <http://dx.doi.org/10.1016/j.chb.2007.02.015>。