

Evaluation of an intelligent tutoring system in pathology: Effects of external representation on performance gains, metacognition, and acceptance.

Crowley, Rebecca S., et al. *Journal of the American Medical Informatics Association* 14.2 (2007): 182-190.

1. Abstract

● Objective

- コンピュータベースチューターリングの効果を発見.
 - ◇ 診断パフォーマンスの向上, メタ認知, 2つの異なる問題表現を使うことの容認
- 課題遂行に要求される診断スキルへのチューターリングの影響を述べる.

※診断：表示された画像の病気を判別すること

1. 診断するために, 画像に含まれる病気の特徴を見つけ出す.
(水疱がある, 腫れている)
2. 見つけた特徴を持つ病気を診断する.
(ミュシャ - ハーバーマン病, 多形性紅斑)

先に病気を推測して, その病気の特徴を探すという診断もある。(鑑別診断)
パターン：特徴の組み合わせのことをパターンと呼ぶ.

● Design

- 問題表現の外在化が異なる2つのインターフェースのどちらかを4時間使う.
 - ◇ 事例焦点表現(the case-focused representation)
 - ◇ 知識焦点表現(the knowledge-focused representation)

● Measurements

- プレテスト, ポストテスト, 保持テスト(retention-test)を実施.
 - ◇ 事例診断テスト
 - ◇ 学生が疑いなく報告したパフォーマンスの比
 - ◇ 参加者アンケート
 - ◇ 学習曲線
 - ◇ チューターリング中のやり取りの行動

● Results

- 一つのチューターリングセッションの後に, 学生は有意に高い学習ゲインを得た.

- 学習は1週間持続した.
- ポストテストもしくは保持テストで, 2つのインターフェースに学習ゲインの違いはなかった.
- 知識焦点インターフェースの学生だけ, プレテストからポストテスト, プレテストから保持テストで, 有意なメタ認知のゲインを示した.
- 学生は, 知識焦点インターフェースを, 事例焦点インターフェースよりも有意に高く評価した.

● Conclusion

- **cognitive tutoring** は, 複合医療領域で改善された診断パフォーマンスと関連している.
- 効果は1週間のポストトレーニングで維持された.
- 知識焦点インターフェースは, メタ認知効果とユーザーの容認で, 事例焦点以上の利点を示した.

2. Background

● Intelligent Tutoring System

- **Intelligent Tutoring System (ITS)** は教育システムで, 1対1指導でよく知られている利益をエミュレートすることを試みる. (Koedinger et al., 1997)
- **Cognitive Tutoring System** は ITS のサブタイプの一つで, 領域特有のプロダクションルールを組み入れている. (Anderson, 1993)
- 多くの **Cognitive Tutoring System** は, たくさんの宣言的知識を必要としない, 手続き的な領域のために開発された.
 - ◇ 数学・科学教育(Koedinger et al., 1993; Gerther et al., 2000)
 - ◇ フライトシミュレータトレーニング(Mitchell, 2000)
 - ◇ 職場でのトレーニング(Lesgold et al., 1993)

● Medical Tutoring Systems

- 本研究以前にも, 少数の医療の ITS が開発されていた. (Clancey & GUIFON, 1983; Eliot et al., 1996; Woo et al., 2006; Azevedo et al., 1998; Sharples et al., 2000; Smith et al., 1998; Suebnukarn & Haddawy, 2006; Kabanza et al., 2006)
- しかしながら, 我々の知る限り, これらのシステムのいずれも, 診断タスクのパフォーマンスを向上させるかどうかを判定するために試験されていない.

● Effect of Problem Representation on Problem-solving

- 認知科学の研究より図的表現が, ロジックを含む領域の数や仮説演繹推論の学習

に関連することが発見された。(Stenning et al., 1995; Toth et al., 2002)

- **Mental Representation for Medical Problem-solving**

- 医療分野の問題解決で2種類の問題表現に関する研究がある。
 - ◇ Case-based(Shmidt et al., 1990)
 - ◇ Knowledge-based(Mandin et al., 1997)
- 私たちの知る限り, case-based 対 knowledge-based の問題表現に関する調査はない。
 - ◇ スキル獲得, メタ認知, ITS による学生の経験について

3. Research Question

- (1) Cognitive tutoring は, 複雑な医療領域での診断パフォーマンスを向上させるのか?
- (2) もし cognitive tutoring が診断パフォーマンスを向上させるならば, その効果はトレーニング期間を過ぎて維持されるか?
- (3) 学生が正しい診断を知っている・知っていないとき, Cognitive tutoring は正しく判断するための学生のメタ認知能力を向上させるか?
- (4) チューターリングシステムを学生が使うことによって, 問題表現(case-focused vs. knowledge-focused)は, 学習ゲインもしくはメタ認知ゲインに作用するか?
- (5) 各問題表現を使うチューターリングシステムの容認で, 学生は異なるか?

4. Method

- **Study Design**

- 同レベルペアのランダム割り当てによって, 被験者のトレーニングレベルの統制を見込み, 本研究では, 被験者間デザインを使った。
- 実験はすべて研究室内で行われた。
- 実験は以下の順で進められる(Figure 1).
 - (1) プレテスト
 - (2) インターフェーストレーニング期間
 - (3) 4.5 時間作業期間
 - (4) ポストテスト
 - (5) アンケート調査
- 1週間後に保持テストを行った。
- すべての参加者は同じ金額を受け取った。

- 作業期間中は, 課題時間は一定にされた。
 - ◇ 20 の皮膚病理学の事例 (表皮下水泡性皮膚炎についての事例)
 - ◇ 20 の事例の順序はすべての参加者で同一

- ◇ 時間内に作業が終わっても、作業期間が終わるまで、初めから再度事例を順に取り組む

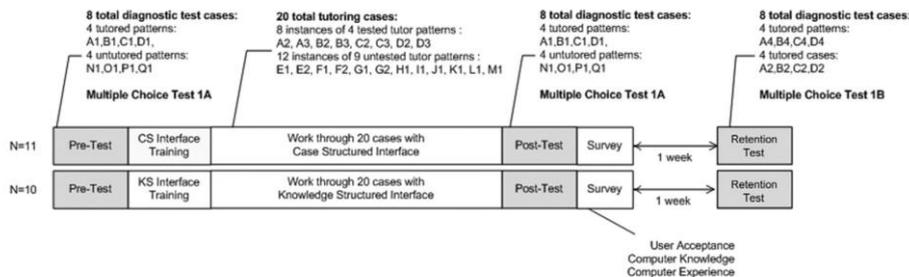


Figure 1. Study design.

● Material

(1) Cases

- 皮膚病理学サービスの活動で、5大学ベースの病理学部の slide archive から事例は得られた。
 - ◇ slide：顕微鏡のスライド板
- 28の表皮下水疱性皮膚炎が選ばれた。
 - ◇ 13の異なる視覚パターン（根拠の組み合わせ）と21の異なる診断の実例
- 20事例は作業期間で使われ、8事例はプレテスト・ポストテストで使われた。(Table 2)
- 同じパターンの複数の事例が、評価（プレ・ポスト・保持テスト）と作業期間の両方で必要不可欠だった。
- 加えて、表皮下水疱性皮膚炎ではない4つの事例を私たちは選択した。
 - ◇ 学習評価に使うための統制事例として
- すべての事例において、診断は他の病理学者によって確認された。

Table 2 ■ Cases Used in Tutoring Sessions, Pretest, Post-test, and Retention Test

Pattern	Diagnosis or Differential Diagnosis	Pattern Appeared in All Tests	Case Appeared in Retention Test
C	Lichen sclerosus et atrophicus	X	
D	Blister above scar	X	
H	Erythema multiforme, Mucha-Habermann disease		
K	Erythema multiforme, Mucha-Habermann disease		
A	Bullous pemphigoid, herpes gestationis	X	X
B	Epidermolysis bullosa acquired, dermatitis herpetiformis, dermatitis herpetiformis-like drug eruption, linear IgA dermatosis	X	
E	Septic vasculitis		X
F	Dermatitis herpetiformis, dermatitis herpetiformis-like drug eruption, linear IgA dermatosis		
L	Spider bite		
M	Epidermolysis bullosa acquired, bullous pemphigoid, herpes gestationis		
G	Porphyria cutanea tarda, variegata porphyria, erythropoietic protoporphyria, hepatoerythropoietic protoporphyria		X
J	Epidermolysis bullosa dermolytic, Epidermolysis bullosa junctional, epidermolysis bullosa acquired, Bart's syndrome		X
D	Blister above scar	X	
E	Septic vasculitis		
B	Epidermolysis bullosa acquired, dermatitis herpetiformis, dermatitis herpetiformis-like drug eruption, linear IgA dermatosis	X	
F	Dermatitis herpetiformis, dermatitis herpetiformis-like drug eruption, linear IgA dermatosis		
A	Bullous pemphigoid, herpes gestationis	X	
C	Lichen sclerosus et atrophicus	X	
G	Porphyria cutanea tarda, variegata porphyria, erythropoietic protoporphyria, hepatoerythropoietic protoporphyria		
I	Dermatitis herpetiformis, dermatitis herpetiformis-like drug eruption, linear IgA dermatosis		

(2) System

- Slide Tutor は皮膚病理学のインテリジェントチューターリングシステムである。
- 視覚分類問題解決のチューターリングのための、私たちの概略の VCT フレームワークの実例の一つである。
- 学生からのヒント要求を受け付け、コンピュータで計算された最善次手を提供する。
- 学生の全活動はキャプチャされて、データベースに保存される。

(3) Interfaces

- 2つの異なるインターフェースを開発した。(Figure 2)

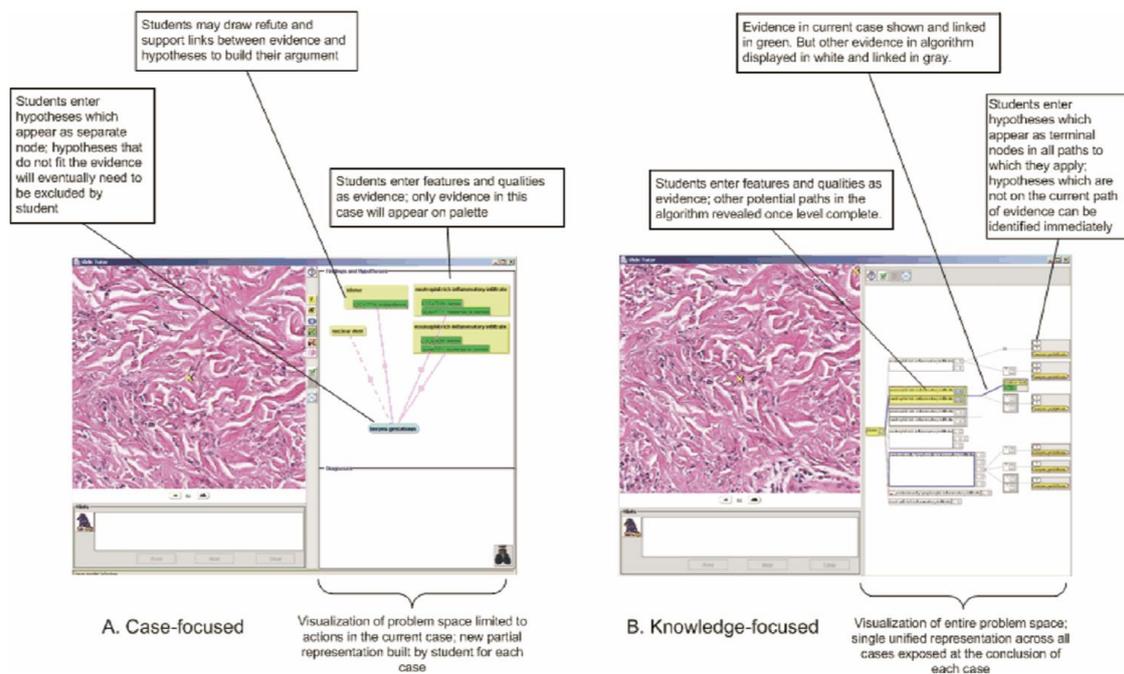


Figure 2. External representations used in (A) case-focused and (B) knowledge-focused interfaces.

- 事例焦点インターフェースは、中心的な焦点として事例を使う問題を示す。
 - ◇ 一つの事例に的を絞った、特徴と推測の関係を表示
- 事例に存在しているもしくは欠如している特徴を学生が確定させたとき、特徴は四角い箱として表示される。
 - ◇ 例えば、Figure 2A のインターフェース左の画像に「水疱がある」という特徴を確定したときに、インターフェース右側の黄色い箱で表示される。
- 学生は、特徴ボックス内に、特徴の箇所や量を記述する。
 - ◇ Figure 2A の黄色い四角内の緑の四角部分
- 推測（の診断）が主張されたとき、それはつながりのないノードとして表示され

る.

◇ Figure 2A の青い四角

- 学生は、特徴と推測の関係を、支持・不支持のリンクで表現できる.
- 診断ができたとき、推測は診断エリアに移されるかもしれない。(DSG の状態と学生のモデル次第である)

◇ Figure 2A の右下部分

- 実際の場合にのみ存在する特徴が表現されているが、任意の有効な推測を添加してテストすることができる.
- 一つずつの事例の推測と特徴の関係を見つつ、学生は 20 事例の関係を構築しなければいけない.

- 一方、知識焦点インターフェースは、全体の知識空間の状況での問題を示す.

◇ すべての事例にまたがった、特徴と推測の関係を表示

- 問題表現はアルゴリズムである.
 - ◇ 問題中の学生の作業として診断ツリーが広がる
- 特徴は四角い箱として診断パス内に表示される.
- 推測は、アルゴリズムの終わりに丸みのある箱で表示される.
- アルゴリズムの残りで見分けるために、正しいパス内の特徴と推測は黄色で示される.
- 推測が現在の根拠と合うときは、推測は現在のパスにつなげられて表示される.
- 問題解決の終わりまで、全体の診断ツリーは診断のために役に立つ.

● Participants

- 21 名の病理学系の学生.
- 大学院の入学年に従って、学生を 2 群に分ける.
 - ◇ 訓練のレベルを均等にするための統制として
 - ◇ 事例焦点インターフェースを使う群：11 人
 - ◇ 知識焦点インターフェースを使う群：10 人
- 参加者には謝金が支払われた.

● Participant Survey

- 参加者は、初日の終わりに 3 つのアンケートを受けた.
- 1 つ目は、25 個の記述について 4 段階で評定.
 - ◇ 記述は、*use of the tutoring system, including enjoyment, future-use, ease of use, specific system features, comparison to alternative methods of learning, self-assessment of learning, trust in content* に関連するもの

- 2つ目, 3つ目はコンピュータ使用とコンピュータ知識を測定
- ◇ 標準化された手段を再現

● Learning Assessments

- プレテスト, ポストテスト, 保持テストはコンピュータ上で行われた.
- それぞれのテストは2つのセクションを含んでいる.
 - ◇ Case Diagnosis Test
 - 参加者は8つの未知の仮想スライド板を詳しく調べた.
 - 参加者は診断もしくは鑑別診断(diagnosis or differential diagnosis)と診断理由を入力する.
 - 評価に使われる事例は以下の3種類がある.
 - A) Untutored Patterns
 - インターフェースを使った作業期間中に見ることのないパターン
 - B) Tutored Patterns
 - 作業期間中に見たことのない事例だが, 作業期間中の別の事例で見たことがあるパターン
 - C) Tutored Cases
 - 作業期間中に見たことがある事例
 - プレ・ポストでは, (4つの事例 A) + (4つの事例 B), 保持テストでは, (4つの事例 B+4つの事例 C) を使った.
 - ◇ Multiple Choice Test
 - 参加者は51項目の多肢選択テストを行った.
 - 特徴を見つけ出す能力, 特徴を確定する能力, 特徴を限定して定量化する能力, 推測に根拠を関係させる能力, 推測の違いを見分けて特徴を選択する能力を評価するために行われた.
 - 作業期間に含まれない表皮下水疱性皮膚炎の領域すべてをカバーしたテスト.
 - プレ・ポストは同じ内容のテスト.
 - 保持テストは別の内容のテスト.

● Scoring of Learning Assessments

- Multiple Choice Test は正解・不正解を皮膚病理学の主治医の回答を基準に採点.
- Case Diagnosis Test の各事例については, 診断と診断理由を個別に採点した.
 - ◇ 診断の採点については,
 - 5点: 最初の正しい診断
 - 3点: (鑑別診断が要求される事例は) 追加の正しい診断一つにつき
 - -1点: 間違った診断一つにつき

- 0点：空欄の回答
- ◇ 診断理由の採点については、
 - 2点：正しい特徴一つにつき
 - -0.5点：間違った特徴一つにつき

● Metacognitive Measures

- すべての Case diagnosis test における各事例の後に、参加者は確かさの点数をつける。
- ◇ 診断が正しいかどうかについて
- ◇ 4段階のリッカート尺度を使用

● Tutoring Process Measures

- 作業期間中のインタラクションデータを収集
 - ◇ マウスクリック、メニュー選択、正しい活動、エラー、ヒント要求
- データ収集の方法は以前記した。(Medvedeva et al,2005)
- データは学習曲線を構成する、表現条件間の処理の違いを説明するために使われた。

● Analysis

- 評価のパフォーマンスは、反復測定分散分析(repeated measures ANOVA)が使われた。
- 主効果と交互作用は、テスト、インターフェース条件、トレーニングレベル、の繰り返しの対比を含む。
- パフォーマンス—確かさの相関とアンケート結果には、t 検定を使った。
- すべての分析は SPSS で行った。

5. Results

● Task Metrics

- 作業期間中に完了できた事例数は、条件間で有意な差はなかった。
- 21人中18人は作業期間中にすべての事例を見た。

● Learning Outcomes

- 両条件とも、multiple choice test と case diagnosis test については、tutored pattern はポストテストで優位に改善された。そして、学習ゲインも維持された。(Table 3)
 - ◇ Multiple choice test の平均スコアはプレテストの 52.8%からポストテストの

- 77.0%に増加した。(MANOVA, effect of test, $F=78.0$, $p<0.0001$)
- ◇ Case diagnosis test の平均スコアはプレテストの 11.7%からポストテストの 50.2%に増加した。(MANOVA, effect of test, $F=64.0$, $p<0.0001$)
 - ◇ Untutored pattern については改善が見られなかった。
 - ◇ 1 週間後の保持テストにおける学習ゲインは保持され、ポストテストと比較しても有意な差はなかった。
 - ◇ 条件間で学習ゲインの有意な差は見られなかった。

Table 3 ■ Pretest, Post-test and Retention Test Scores

Condition	Pretest			Post-test			Retention Test		
	Multiple Choice	Case Diagnosis		Multiple Choice	Case Diagnosis		Multiple Choice	Case Diagnosis	
		Tutored Patterns	Untutored Patterns		Tutored Patterns	Untutored Patterns		Tutored Patterns	Tutored Cases
Combined	53 ± 13	12 ± 9	12 ± 8	77 ± 10	50 ± 23	10 ± 7	74 ± 13	48 ± 21	34 ± 17
CF	53 ± 13	12 ± 8	12 ± 9	76 ± 11	51 ± 25	8 ± 6	75 ± 14	45 ± 22	28 ± 19
KF	53 ± 13	13 ± 10	12 ± 7	79 ± 11	50 ± 21	13 ± 7	74 ± 14	51 ± 20	41 ± 13

MANOVA, repeated contrast, $F=78.0$, $p<0.001$; MANOVA, repeated contrast, $F=3.52$, $p=0.08$; MANOVA, repeated contrast, $F=64.01$, $p<0.001$; MANOVA, repeated contrast, $F=0.25$, $p<0.62$; MANOVA, repeated contrast, $F=0.89$, $p=0.36$

● Metacognitive Measures

- 確かさーパフォーマンスの相関は、自身の知識や能力を評価するための学生の能力を測定する。
 - ◇ ポストテスト、保持テストで有意に正の相関があった。
 - ◇ 加えて、プレテストからポストテストへの、case diagnosis test のスコアの変化と確かさの変化の間に正の相関があった。(r=0.46, p=0.04)
- 回帰分析によって、各テスト、各条件の確かさとパフォーマンスの傾きを求めた。
 - ◇ この分析は知識焦点条件の有利な点を明らかにした。
 - ◇ 確かさーパフォーマンスの傾きは、プレからポストで有意に増大した。(p<0.05)
 - ◇ 確かさーパフォーマンスの傾きは、ポストから保持テストで有意に増大した。(p<0.01)
 - ◇ この効果は知識焦点条件のみに見られた。

● Participants Survey

- アンケートの結果は知識焦点条件の方が有意に高かった。(t=2.66, p=0.02)
- 特に、以下のものは知識焦点条件の方が高かった。(Figure 3)
 - ◇ Future Use (p=0.001)
 - ◇ Enjoyment (p=0.05)
- どちらの条件でも、アンケート結果と学習結果の間に相関は見られなかった。

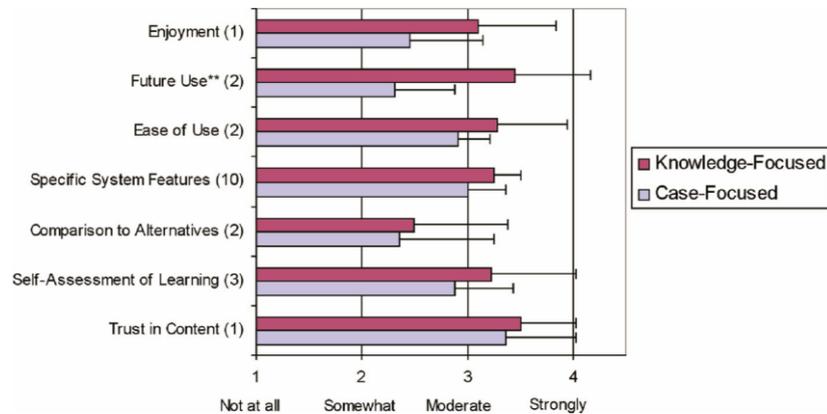


Figure 3. Survey results.

● Process Measures

- 分析は、正しい活動，エラー，ヒント要求に焦点を当てている。
 - ◇ ユーザーはどこでヒントを要求するのか
 - ◇ 何種類のエラーが共通しているのか
 - ◇ どうやってヒントを使ってエラーを変えるのか
- 両条件ともヒントの使用は一定のレベルで留まっている。(Figure 4)
- 両条件ともエラーは急速に減少する。

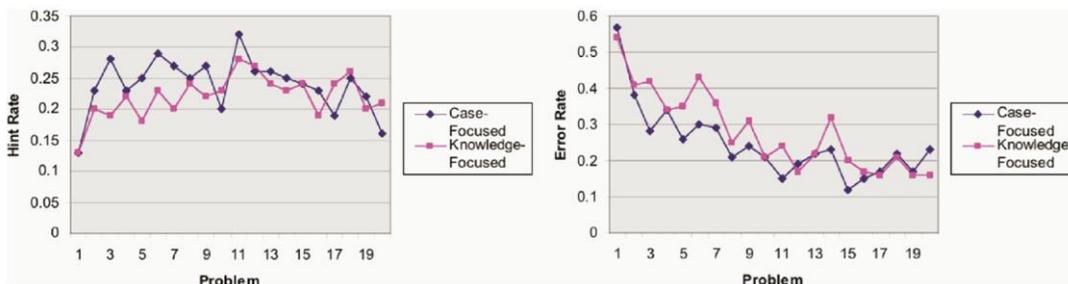


Figure 4. Hint and error rates over time by condition.

6. Discussion

● Significance

- 医療分野における cognitive tutoring の広範囲による評価はこれが初めて。
- 1週間空いても学習ゲインは変わらずに維持された。
- どちらの条件も tutored pattern についての学習ゲインは獲得した。
- どちらの条件も untutored pattern についての学習ゲインは得られなかった。

● Limitations of Present Study

- サンプルサイズ (N=21) が小さかった。
- この研究では、cognitive tutoring パラダイムを標準的な手法に対してテストしていない。