

## Outcome feedback during collaborative learning: Contingencies between feedback and dyad composition

Christa S.C. Asterhan, Baruch B. Schwarz, Noa Cohen-Eliyahu

Learning and Instruction vol.34 1-10(2014)

### 1. 導入

- フィードバック研究は、個人を対象にした研究は膨大に存在する。
- そのため、フィードバックと協同ペアの構成の影響について調査する。

#### 1.1 フィードバックと協同

- Krause, Stark, and Mandl(2009)では、結果のフィードバックを受けることで、パフォーマンス改善が報告された。
  - 単独条件, ペア条件のどちらでも
- Schwarz and Linchevski(2007)では、中学3年生の推論タスクでのパフォーマンスを調査。
  - フィードバックあり, ペア条件 → パフォーマンスの向上あり。
  - フィードバックなし, 単独条件 → パフォーマンスの改善なし。
  - この研究ではフィードバックと協同の独立した効果の検討ができていない。
- Ellis, Klahr, and Siegler(1993)では、小学5年生が小数のために使う数学的規則について、結果のフィードバックと協同の効果を検討した。
  - フィードバックあり, 協同条件 → 優れた学習利益が得られる。
- Tudge らによる二つの研究では、同様に小学生の数学的推論に焦点を当てている。
  - いかなる結果のフィードバックも受けないとき (Tudge ら, 1996)
    - ◇ 協同の条件 > 単独の条件, 学習の効果。
  - 単独条件でも、フィードバックを受ける、利点がある (Tudge & Winterhoff, 1993a)

#### 1.2 協同学習におけるペアの効果

- ペアの構成は、それぞれのペアが問題に対して、どのような知識を有しているかで定義。
  - 間違い知識-正しい知識ペア (W-R)
  - 間違い知識-間違い知識ペア (W-W)
- Vygotskian 学者によると、より有能なペアとの相互作用はより良い学習につながる。
- Neo-Piagetian 学者によると、間違っ異なる初期認知 (Wx-Wy) を持つペア間のやりとりにより焦点を当てている。
  - いくつかの研究では、Wx-Wy ペアのやりとりの方が W-R ペアのやりとりよりも利益がある (Ames & Murray, 1982; Doise & Mugny, 1979; Doise et al.,

1975;Glachan&Light,1982;Schwarz et al.,2000)

- Ames and Murray(1982)
  - Wx-Wy ペアにおいて、対話の機会が与えられている場合にのみ成長が示された。

### 1.3 結果のフィードバックとペアの合併効果

- フィードバックの効果はペア構成に依存していることを発見した2つの先行研究(Ellis et al., 1993;Tudge et al.,1996)
  - 間違い知識-間違い知識ペア (W-W)
  - Tudge らの研究では「より有能な」ペアとは、間違っているがより一層洗練された説明を提供したパートナーかもしれない。(Wx-Wy)
- 間違い知識-正しい知識ペア (W-R) がフィードバックを受けて議論を行うことで
  - W を有する学生は、R 有する学生の持つ高度な解決方略にさらされる
  - R の方略が正しい解答を導くことの経験的な確認も得られる
  - この組み合わせは非常に強力な学習効果が予想される
- 間違い知識-間違い知識ペア (W-W) がフィードバックを受けて議論を行うことで
  - Wx-Wx 同士では議論ですぐにお互いの解答に同意できることが予想できる。
  - しかし、負のフィードバックが返ってくるため、外部の支援なしで新しい方略を導かなければならない。

### 1.4 仮説

- 仮説 1 :
  - 結果のフィードバックが学習の利益を向上させる。
- 仮説 2 :
  - 別の解答を提案してくる学生 (Wx-Wy ペア) でのやりとりは、単独条件や同じ能力のペア条件 (Wx-Wx ペア) でのやりとりよりも学習に有益。
- 仮説 3 :
  - R とペアの状態 (Wx-R ペア) でのやりとりは、単独条件や同じ能力のペア条件 (Wx-Wx ペア) よりも学習に有益。
- 仮説 4a :
  - Wx-Wy ペアでのやりとりは、Wx-R ペアでのやり取りよりも学習に有益。
- 仮説 4b :
  - Wx-R ペアでのやりとりは、Wx-Wy ペアでのやり取りよりも学習に有益。
- 仮説 5 :
  - 結果のフィードバックとペア条件間の交互作用がある。

## 2.手法

### 2.1 参加者

- エルサレムとテルアビブ首都圏（イスラエル）の 9 つの公立中学校の、中学 3 年生のクラスを対象にしている.
- 9 つの学校から合わせて 600 名の 14~15 歳の学生.
- プレテストで比例推論タスク (proportional reasoning tasks) 行う
  - 各学生が問題を解くために使う方略を評価するため
  - プレテストの結果, 104 人の学生が実験の対象から除外された.
    - ◇ この実験では, 学生自身の解答について説明することが求められる.
    - ◇ しかし, 104 人の学生は自身の解答について説明できるレベルに達していないため, 除外された.
- 残った学生 496 人 (男: 301 人, 女: 195 人) を, 学生が使う方略で分類分けを行う.
  - 加法推論方略 (additive reasoning strategy) : W1, N=196
  - 初期的な比例推論方略 (proto-proportional reasoning strategy) : W2, N=194
  - 比例推論方略 (proportional strategy) : R, N=106
- 介入タスク (intervention task) に参加.
  - 学生はランダムでペア条件を割り当てられた.
  - 72 人の R 学生は W1, W2 学生とペアを組んだ.
  - 残り 34 人の R 学生は別のタスクに参加. (分析対象外)

### 2.2 計画

- プレテストのパフォーマンスに基づいて, 学生を W1, W2, R に分類し, 5 つのペアオプションから 1 つのペア, もしくは, 単独条件へランダムに割り当てを行った. (Table 1 参照)
  - ペアオプション
    - ◇ W1-W1, W1-W2, W2-W2, W1-R, W2-R
  - 単独
- 同質の方略を持つペアは  $W_x-W_x$  と表現する. (W1-W1 もしくは W2-W2)
- 異質の方略を持つペアは  $W_x-W_y$  と表現する. (W1-W2)
- R の方略を持つ学生とのペアは  $W_x-R$  と表現する. (W1-R もしくは W2-R)
- 学生は, 介入タスクで結果のフィードバックが得られるかどうかランダムで決定.
- 実験計画:  $2 \times 4$ 
  - 結果のフィードバック条件: フィードバックあり, フィードバックなし
  - ペア条件: 単独,  $W_x-W_x$ ,  $W_x-W_y$ ,  $W_x-R$

### 2.3 道具

- プレテスト, 介入タスク, ポストテストで使われるタスクはすべてブロックタスク.
  - ブロックタスクは, Harel ら(1992)によって, 学生を数学能力で区別する診断目的で設計された.
- 学生は 4 つの 3 次元構築物 (A, B, C, D) が示される.
  - これらの構築物はいくつかのブロックから構成されている.
  - A と C を構成するブロックは同じ色で示され, 同じ重さのブロックである.
  - 同様に, B と D を構成するブロックは同じ色で示され, 同じ重さのブロックである.
  - 各試行, 学生には最初に 2 つの構築物 A と B の関係について情報が与えられる.  
(A は B よりも重い, B は A よりも重い, A と B は同じ重さ)
  - そして, 学生は 2 つのターゲット構築物 C と D の関係を解答する. 解答は 4 つの選択肢から一つ選ぶ.  
(C は D よりも重い, C は D よりも軽い, C と D は同じ重さ, これらの情報だけでは決められない)
- 学生は, 解答を選択した後に, なぜその解答を選択したのか言語説明が求められる.  
(Fig.1 参照)

### 2.4 5 つのブロックタスク

- このタスクは筆記テストで行われる.
- 学生の推論方略を評価するために 5 つの難易度のアイテムが用意されている. (Table 2 参照)
- アイテムは, 異なる推論方略を使うことで, 異なる解答が導かれるように設計されている.
  - 例えば Tabel 2 の Item 3 において,
    - 1 プレテストで除外された 104 人は, C と D のブロックの個数から解答を導く.
    - 2 W1 の学生は, A と B の重さが等しく, C は A より 4 つブロックが多く, D も B より 4 つブロックが多いので, C と D の重さは等しいという解答を導く.
    - 3 W2 の学生は, A は 26 個のブロックで B の 27 個のブロックの重さと等しくなる. これは A のブロックの方が, B のブロックよりも重いことを意味している. そして, C は A より 4 つブロックが多く, D も B より 4 つブロックが多いので, C の方が D よりも重いという解答を導く.
    - 4 R の学生は, A のブロックの重さを  $a$ , B のブロックの重さを  $b$  とすると,  $26a=27b$ .  $a=(27/26)*b$  より  $30a=31.15b$ .  $30a$  (C の重さ)  $>31b$  (D の重さ) より C の方が D よりも重いという解答を導く.
- W2 と R はほとんど同じような推論方略だが, Item4 においては, W2 は「これらの情

報だけでは決められない」という間違っただけの解答を導く。R は Item4 でも正しい解答を導く。

## 2.5 介入タスク

- =学習タスク
- 単独条件，二人組条件のいずれかにランダムで割り当てる
- 2つの追加ブロックタスク (I1, I2) を解く。(Table 3 参照)
  - このタスクでは，ブロックの物理的な模型と計量器が学生に提供される。
  - I1, I2 は W1, W2 の方略では誤った解答を導く問題になっている。

## 2.6 符号化処理

- それぞれのテストアイテム (プレテストの 5 つ，介入タスクの 2 つ，ポストテストの 5 つ) を解くために，学生がどの方略を使ったか，それぞれのアイテムごとに以下の 4 つのカテゴリに分類する。(Table 4 参照)

grade 1：視覚推論方略

grade 2：加法推論方略

grade 3：初期的な比例推論方略

grade 4：比例推論方略

- 二人の独立した評価者によって符号化された (評価者間の信頼性：Cohen's  $k=0.925$ )
- 学生の方略レベルは以下の基準に従う。
  - W1：5つのそれぞれプレテストアイテムに加法推論方略を使った。
  - W2：最低 1 回，初期的な比例推論方略を使った。しかし，比例推論方略は使っていない。
  - R：最低 1 回，比例推論方略を使った。
- プレテスト，ポストテストのパフォーマンスは，それぞれのテストの 5 タスクアイテムの平均 grade から計算される。

## 3.手順

- Stage 1 プレテスト
  - プレテスト 5 問
  - 教室の前方にアイテムの模型 (A, B, C, D) を展示
  - 25~40 分 (学生が 5 問終わらせるのにかった時間)
- Stage 2 介入タスク
  - 介入タスク 2 問
  - 単独条件 もしくは ペア条件
  - 結果のフィードバックあり，なし

- 教室の前方にアイテムの模型 (A, B, C, D) を展示
- Stage 3 ポストテスト
  - ポストテスト 5問
  - 教室の前方にアイテムの模型 (A, B, C, D) を展示
  - Stage 2 と Stage 3 のインターバルは 2~3 週間

#### 4.結果

- プレテスト, ポストテストの結果が Table 5 に示される.
- 分析は混合モデル (SAS PROC MIXED) で行われた.
  - 変量効果 dyad within condition, individual within dyad and condition
  - RQ を調査するために, 2つの区別されたモデルを構成.
  - 外れ値はトリミングされた.

##### 4.1 全ての協同の効果と結果のフィードバック

- フィードバックあり・なし, 単独・ペアで結果をまとめた表が Table 6 に示される.
- 結果のフィードバックに有意傾向あり.  $F(1,358)=3.78, p=.053, R^2=.04$ 
  - 仮説 1 に一致
- 協同の主効果, 2 要因間の交互作用なし.

##### 4.2 ペア条件の特異的効果と結果のフィードバック

- ペア条件 (単独, 同質ペア [W<sub>x</sub>-W<sub>x</sub>], 異質ペア [W<sub>x</sub>-W<sub>y</sub>], R ペア [W<sub>x</sub>-R]) × フィードバック条件 (あり, なし) の 8 条件の平均ゲインスコアを Fig.2 に示される.
- 結果のフィードバックに主効果はない.  $F(1,236)=2.53, p=.113$
- ペア条件に主効果が見られた.  $F(2,231)=12.31, p<0.0001, R^2=.21$
- 事後分析 (Tukey-Kramer adjustments) は仮説 3 に一致したことを示す.
- R とのペアを組んだ学生は, W<sub>x</sub>-W<sub>x</sub> ペアと比べてより大きなゲインが生じた.  $t(223)=6.08, p<.001$
- R とのペアを組んだ学生は, 単独と比べてより大きなゲインが生じた.  $t(280)=2.94, p=.019$ 
  - 仮説 4a は矛盾したが, 仮説 4b は確認された.
- R とのペアを組んだ学生は, W<sub>x</sub>-W<sub>y</sub> ペアと比べてより大きなゲインが生じた.  $t(215)=3.05, p=.014$ 
  - 仮説 2 は矛盾した.
- W<sub>x</sub>-W<sub>y</sub> ペアを組んだ学生, W<sub>x</sub>-W<sub>x</sub> ペアよりも大きなゲインは見られなかった.  $t(191)2.31, p=.010$
- W<sub>x</sub>-W<sub>y</sub> ペアを組んだ学生, 単独よりも大きなゲインは見られなかった.  $t<1$

- ペア条件と結果のフィードバック間で有意な交互作用が見られた。  $F(2,231)=3.26$ ,  $p=.022$ ,  $R^2=.06$ 
  - 仮説 5 と一致した。
- 多重比較のための Tukey-Kramer テストは、次の傾向を明らかにする。
  - フィードバックあり条件では、R とのペアを組んだ学生は、 $W_x-W_x$  ペアと比べてより大きなゲインが生じた。  $p<.001$
  - フィードバックあり条件では、R とのペアを組んだ学生は、 $W_x-W_y$  ペアと比べてより大きなゲインが生じた。  $p<.001$
  - フィードバックあり条件では、R とのペアを組んだ学生は、単独と比べてより大きなゲインが生じた。  $p=.010$
  - フィードバックなし条件では、R とのペアを組んだ学生は、 $W_x-W_x$  ペアと比べてより大きなゲインが生じなかった。  $p=.082$
  - フィードバックなし条件では、R とのペアを組んだ学生は、 $W_x-W_y$  ペアと比べてより大きなゲインが生じなかった。  $t<1$
  - フィードバックなし条件では、R とのペアを組んだ学生は、単独と比べてより大きなゲインが生じなかった。  $t<1$
  - フィードバックあり・なし条件で、与えられたペアの条件を含んだ比較は次の傾向を示す。
    - R とのペア組み、フィードバックがあったときだけ、よりゲインが大きかった。  $t(232)=3.40$ ,  $p=.018$
- 結果のフィードバックがあるだけでも、R とペアを組んでいるだけでも学習ゲインは生じず、結果のフィードバックがあり、かつ、R とペアを組んでいるときだけである。
  - これは事後比較の発見によって支援される。
- 結果のフィードバックがあり、かつ、R とペアを組んでいる学習者は、統計上有意に高いゲインのスコアを示した。
  - ほかの 7 条件ではこのスコアへ届かなかった。

## 5 考察

- ペアの組み合わせを考慮しなかったとき、単独・ペアで、結果のフィードバックによる学習利益があった。(仮説 1) ペアの組み合わせを考えない単純な協同は、単独条件と比較した場合に、優れた学習成果は見つからなかった。
- ペアの組み合わせで細分化したとき、学習利益のペア条件の主効果は、 $W_x-R$  条件が他の条件 ( $W_x-W_x$ ,  $W_x-W_y$ , 単独) よりも上回った。(仮説 4b と仮説 3)
- 予想に反して、 $W_x-W_y$  は  $W_x-W_x$  よりも高い学習利益をもたらさなかった。(仮説 2)
- 結果フィードバックの効果は、ペア条件と交互作用がある。(仮説 5) 結果のフィードバックは R とペアを組んだ学生だけに恩恵がある。

### 5.1 協同学習における結果のフィードバック

- 他の協同学習と結果のフィードバックの研究では、矛盾した結果のものがある。これらは、協同グループの構成を考えていない。
- 協同学習における結果フィードバックの役割を検討する際に、グループ構成の違いを特定することの重要性を強調している。
- R とペアを組んだ学生だけが結果フィードバックから利益を得る。しかし、単独条件でも結果のフィードバックがプラスの効果がある。
- $W_x-W_x$ ,  $W_x-W_y$  では、認知的葛藤が生じて、フィードバックの結果を無視しているのかも知れない。
- 今後の研究では、介入タスクの詳細なプロセスデータ収集することで、ペア条件でのフィードバックの無効に焦点を当てるべき。

### 5.2 協同学習におけるペアの組み合わせ

- R とやりとりのある学生は、高次の推論方略にさらされるので、「より有能なペアとの相互作用はより良い学習につながる」というヴィゴツキーの理論を裏付けているだろう。
- $W_x-W_y$  ペアが他の条件に勝らなかったという知見は、新ピアジェ説明の「誤りの誤りは正解」現象と明らかに矛盾している。先行研究でこの現象が成り立つのは  $W_x$  と  $W_y$  が等しく間違っているからだが、本研究の  $W_x$  と  $W_y$  は、 $W_x$  を洗練したものが  $W_y$  のような階層的な方略になっているのでこの現象は成り立たない。

### 5.3 研究の限界と今後

- グループのサイズが小さい。
- 先行研究に沿った参加者の年齢層で調査した。
- 推論方略は、階層的に編成されている。
- 学生は自分の知識の精度や正確さに自信がある。

**Table 1**  
 Distribution of participants to dyadic pairing conditions, based on own and partner's competence level.<sup>a</sup>

Pairing condition		Feedback condition	
Dyad member 1	Dyad member 2	Without feedback	With feedback
W <sub>1</sub>	–	N = 22	N = 15
W <sub>1</sub>	W <sub>1</sub>	N = 40	N = 40
W <sub>1</sub>	W <sub>2</sub>	N = 44	N = 44
W <sub>1</sub>	R	N = 36 <sup>b</sup>	N = 34 <sup>b</sup>
W <sub>2</sub>	–	N = 13	N = 18
W <sub>2</sub>	W <sub>2</sub>	N = 42	N = 40
W <sub>2</sub>	R	N = 30 <sup>b</sup>	N = 44 <sup>b</sup>

<sup>a</sup> W<sub>1</sub> = student who employed additive reasoning strategies only on pretest, W<sub>2</sub> = student who used a proto-proportional strategy at least once, R = student who used a full-fledged, proportional strategy at least once.

<sup>b</sup> Number of participants in these cells include the R partners of W<sub>1</sub> and W<sub>2</sub> participants. Total number of participants, excluding the R partners is N = 390.

**Table 2**  
 The given and the solutions of the five items of the 5BlocksTaskTest.

Item	N(A)	N(B)	N(C)	N(D)	Given weight relation between A and B	Weight relation to be found between C and D
1	27	27	30	29	weight(A) > weight(B)	weight(C) > weight(D)
2	26	27	30	30	weight(A) < weight(B)	Undetermined
3	26	27	30	31	weight(A) = weight(B)	weight(C) > weight(D)
4	10	9	31	28	weight(A) < weight(B)	weight(C) < weight(D)
5	10	16	24	37	weight(A) = weight(B)	weight(C) > weight(D)

**Table 3**  
 The given and the solutions of the two Block task items in the intervention phase.

Item	N(A)	N(B)	N(C)	N(D)	Given weight relation between A and B	Weight relation to be found between C and D
I1	11	10	34	31	weight(A) < weight(B)	weight(C) < weight(D)
I2	10	12	24	27	weight(A) = weight(B)	weight(C) < weight(D)

**Table 4**  
 Coding categories for students' proportional reasoning strategies.

Category	Description	Examples
Visual reasoning	Judgment is based on visual features of blocks C and D or reached by counting the number of bricks in those two blocks. Blocks A and B are ignored.	"C is heavier because it has more bricks than D."
Additive reasoning strategy	The student relates to all four blocks, considering the difference between the number of bricks to establish his/her judgment. Some explanations at this level compare the difference between A and B to that between C and D; others compare the difference between A and C to that between B and D.	"The difference between A and C is that C has 3 more bricks and A is lighter. The difference between B and D is also three bricks more to D, so C will be lighter."
Proto-proportional reasoning strategy	The student relates blocks C and D to blocks A and B, considering the weight of bricks relatively to whole blocks.	"In B there are less bricks than in A, and B weighs more, so each brick in B weighs more than in A. In D there are less bricks than in C. I know that each brick in D weighs more, but there are more bricks in C than in A and in D than in B. So I don't know. I can't decide if this one brick is not too little to balance"
Proportional reasoning strategy	In addition to considering the weight of bricks relative to whole blocks, explanations at this level take into account discrepancies between blocks relative to their magnitude.	"26 bricks of B weigh more than 27 bricks of A. In D there are 29 bricks and in C, 30. I know that 26/27 is less than 29/30, so D will be heavier."

**Table 5**  
 Raw mean (and SD) pretest and posttest scores on the 5BlocksTaskTest,  $N = 390$ .

Own level of competence	Pairing condition	With feedback				No feedback			
		Single	With W <sub>1</sub> peer	With W <sub>2</sub> peer	With R peer	Single	With W <sub>1</sub> peer	With W <sub>2</sub> peer	With R peer
W <sub>1</sub>	Pretest	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)	2.00 (.00)
	Post-test	2.09 (.14)	2.17 (.66)	2.25 (.36)	2.74 (.72)	2.25 (.50)	2.04 (.16)	2.29 (.42)	2.33 (.36)
W <sub>2</sub>	Pretest	2.45 (.28)	2.35 (.22)	2.37 (.21)	2.55 (.27)	2.62 (.33)	2.48 (.27)	2.50 (.28)	2.49 (.37)
	Post-test	2.78 (.52)	2.52 (.59)	2.47 (.41)	3.05 (.46)	2.71 (.59)	2.58 (.48)	2.48 (.52)	2.64 (.53)

**Table 6**  
 Estimates of marginal mean (and SE) learning gains of W students, by collaborative condition (dyadic or singleton) and outcome feedback condition (with or without),  $N = 390$ .

	Singleton	Dyadic	Total
Outcome feedback	.17 (.07)	.24 (.03)	.20 (.04)
No outcome feedback	.16 (.07)	.11 (.03)	.14 (.04)
Total	.17 (.04)	.18 (.02)	

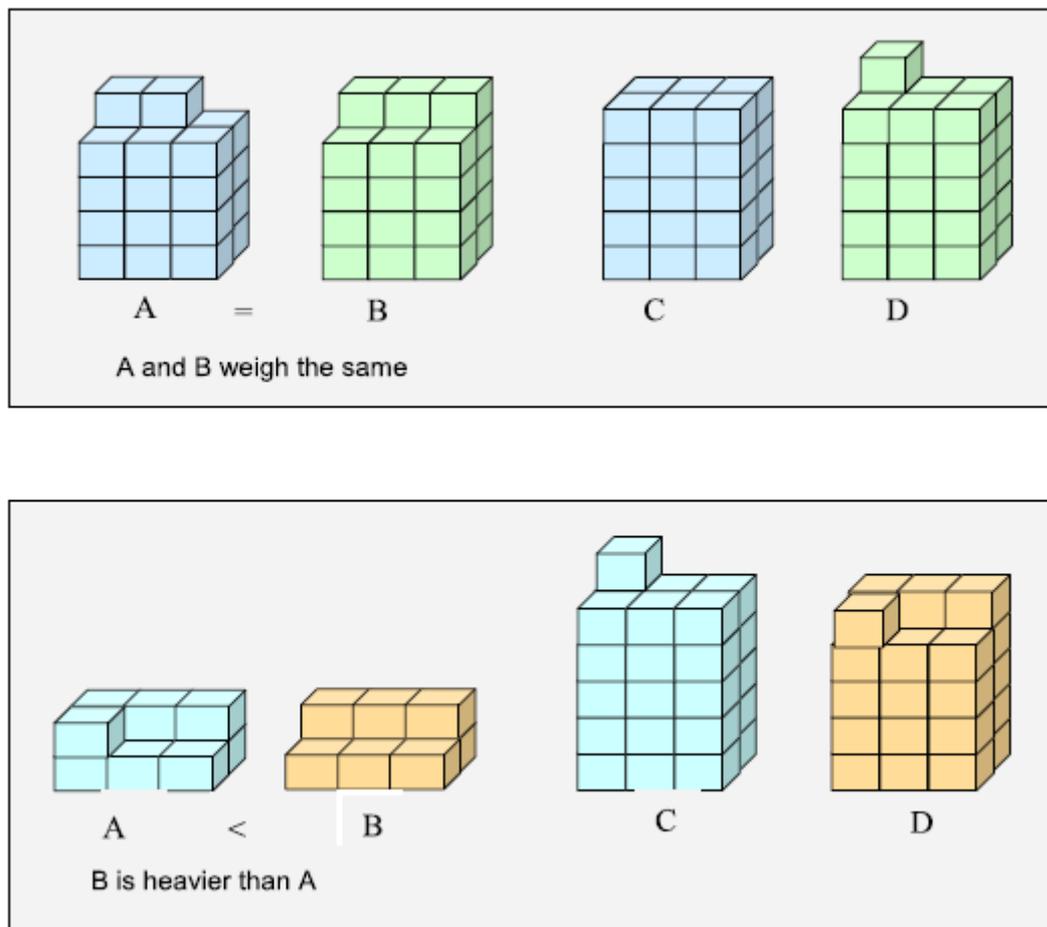


Fig. 1. Two items of the 5BlocksTaskTest (items 3 and 4).

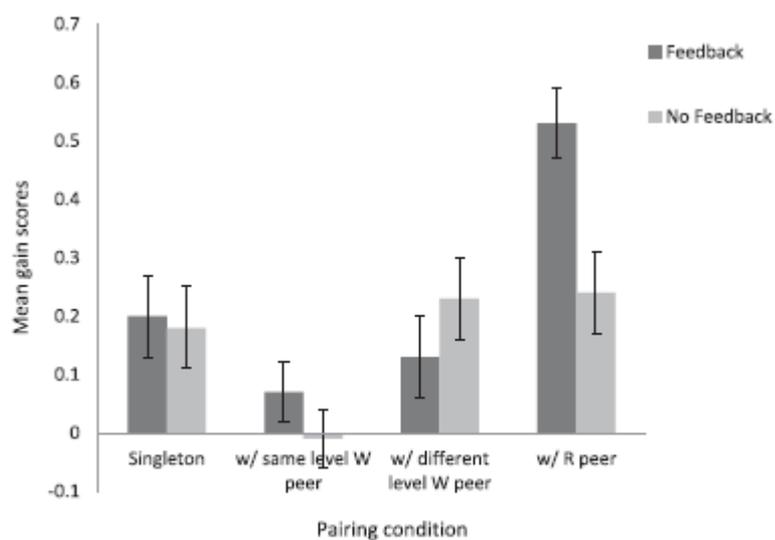


Fig. 2. Estimates of marginal mean (and SE) gain scores for  $W_x$  by pairing and outcome feedback condition.