

Uncovering Expertise-Related Difference in Troubleshooting Performance: Combining Eye Movement and Concurrent Verbal Protocol Data

TAMARA VAN GOG, FRED PAAS and JEROEN J. G. VAN MERRIËNOBER

■ Introduction

- トラブルシューティングにおけるエキスパートとノービスの差異
(Chi, Glaser, & Farr, 1988; Ericsson & Lehmann, 1996)
 - ◇ 知識の多少
 - 領域知識 多
 - システムの心的な表象の構築が容易
 - 効果的な方略を選択できる
 - 領域知識 少
 - 領域に依存しない一般的な方略に頼らざるを得ない
 - ◇ 知識の体系化
 - 長期記憶へのアクセシビリティの良否
 - ◇ 欠陥ベースの推論 / システムベースの推論
 - 欠陥ベース (エキスパート)
 - 良く似た不具合の事例がメンタルモデルに含まれる
 - 類似性に基づいてシステムを診断したり, 修復したりすることができる
 - システムベース (ノービス)
 - システムについての心的表象の構築
 - 表象を用いてシステムの振る舞いを推論
- エキスパートとノービスの差異については多くの研究対象が存在する一方で, ノービスがある領域のエキスパートになるまでの比較的長い期間についてはあまり知られていない
 - ◇ エキスパートとノービスのパフォーマンスの差はこの期間の中で生じる 熟達化
 - 複雑な認知的スキルの場合, この期間は非常に長い
 - ◇ 知識やパフォーマンスの差について研究することで熟達化の研究を発展させることができる
- Process-tracing techniques
 - ◇ 認知的プロセスや知識を推測することが可能
 - 複雑な認知的タスクについて, 熟達度の違いによって生じるパフォーマンスの差を研究するのに非常に有効である
 - ◇ 言語報告 (Ericsson & Simon, 1993)
 - 最も広く使われる process-tracing techniques
 - 同時的言語報告 「声に出して考える」 ('think aloud')
 - ◇ タスクにあたる間ずっと, 思い浮かんだことをすべて声に出す

- 回顧的言語報告
 - ◇ タスク終了直後に, 問題を解いている間どのように考えたかを報告する
 - 言語報告の特徴
 - 認知プロセスの推測に利用できるが, 推測が正しいことは保証されない
 - 教示の言い回しが結果に大きな影響を与える
 - ◇ 眼球運動
 - 言語報告ほど一般的ではない
 - タスクにあたる間の眼球運動を記録
 - 眼球運動の特徴
 - 注意の焦点を知ることができる 認知プロセスの推測 (Rayner, 1998)
 - ◇ 注意のシフト (Rayner, 1998; Stelmach, Campsall, & Herdman, 1997)
 - 外因的
 - ✓ 主に環境内の突出した特徴によって引き起こされる
 - 内因的
 - ✓ タスクや環境, 情報源の重要性についての知識によって引き起こされる
熟達度によって影響をうける
 - 認知的負荷の情報を知ることができる
 - ◇ 停留時間 (Underwood, Jebbett, & Roberts, 2004)
 - 認知的負荷の増大に伴って増大
 - ◇ サッケードの距離
 - 認知的負荷の増大に伴って減少
 - ◇ 眼球運動と同時的言語報告の組み合わせは, 比較的小さな熟達度の違いを検討するために有効である
- Troubleshooting
 - ◇ 本研究ではコンピュータ上でシミュレートされた不具合のある電気回路を扱う
 - ◇ 問題を解決するプロセスは物理的な行動 (マウスやキーボードの操作など) に基づいて以下のフェーズに区別することが可能
 - Phase 1: 'problem orientation'
 - 回路を判断する 心的な表象の構築
 - Phase 2: 'problem formulation and action decision'
 - 問題点を明確にし, 最初の行動を決定する
 - ◇ 不具合の診断を含む
 - Phase 3: 'action evaluation and next action decision'
 - 行動の結果を評価し, 次の行動を決定する
 - ◇ 不具合の診断を含む
 - Phase 4: 'evaluation'
 - 問題に対する解決策の最終的な評価

- ◇ Phase 3 の繰り返し
 - 診断した原因が必ずしも正しくなかったり、複数ある原因の 1 つでしかなかったりするため
 - ◇ 下位の phase
 - Phase 2: problem formulation と action decision
 - Phase 3: action evaluation と next action decision
 - これらは物理的な動作からは区別することはできない
 - ◇ この研究では phase 1 ~ 3 に集中する
- 予測
- ◇ 熟達化期間の初期段階の被験者にタスクを与える場合
 - 予測 1: 熟達度の高い群 (higher 群) は, 低い群 (lower 群) に比べて, problem orientation と problem formulation に多くの時間を費やす
 - Higher 群
 - ◇ メンタルモデルを問題の心的表象と組み合わせて推論に用いることができる
 - ◇ まず心的表象を構築しようとする
 - Lower 群
 - ◇ 心的表象の構築が困難
 - ◇ 回路のテストを行い, そこから得た情報をもとに仮説を形成しようとする
 - 予測 2: Higher 群は, lower 群に比べて, action decision と action evaluation に多くの時間を費やす
 - Higher 群
 - ◇ 取った行動の影響を考える
 - ◇ その行動によりゴールに近づいたかを見積もる
 - Lower 群
 - ◇ 明確なゴールがないまま思い浮かんだものを試す
 - ◇ その結果から新たな仮説を生成しようとする
 - 予測 3: problem orientation と evaluation は problem formulation や action decision よりも認知的な負荷が小さい
 - 予測 4: Higher 群は, lower 群に比べて, すべてのプロセスで認知負荷が小さい
- 仮説
- ◇ 先行研究 (Haider and Frensch, 1999; Charness et al., 2000) 結果に基づく
 - ◇ Problem orientation (フェーズ 1) において
 - Higher 群は停留時間のパーセンテージが高い
 - 停留点は主要な欠陥に関するコンポーネント間で切り替わる

• Method

- ◇ ここで報告するデータはより大きな実験の一部として集められたものである
 - 第 1 の目的
 - 各種の言語報告 (同時的, 回顧的) を眼球運動及びマウス・キーボード操作記録に基づいて比較する
 - 第 2 の目的
 - 眼球運動のデータが, 熟達化が関係するパフォーマンスの違いを調べるために有効かどうかを評価する
 - 眼球運動とプロトコルの関係を調べる

- Design
 - ◇ ラテン方格法により 4 つのシーケンスを被験者内でランダムに割り当て
 - シーケンス 1: 同時報告 : タスク 1, タスク 2
 - シーケンス 2: 同時報告 + 眼球運動 : タスク 3, タスク 4
 - シーケンス 3: 回顧的報告 : タスク 5, タスク 6
 - シーケンス 4: 回顧的報告 + 手がかり : タスク 7, タスク 8
 - ◇ ここではタスク 3 (同時報告 + 眼球運動の 1 問目) の結果を用いる
 - ◇ 課題の内容は共通だが, 課題を解いたタイミングは被験者により異なる

- Participants
 - ◇ pre-university education の 5 年目, または higher professional education の 1~2 年目の学生
 - 少なくとも電気の基礎理論を学んでいる
 - ◇ 全 26 人の被験者から, 全課題の成績の上位と下位 5 人ずつを選出
 - この成績は performance, mental effort, time-on-task を総合したもの (Table 1)
 - ◇ 尺度の詳細は後述
 - Higher 群
 - 全員 higher professional education の学生
 - seq. 1: 1 人; seq. 2: 2 人; seq. 3: 1 人; seq. 4: 1 人
 - Lower 群
 - 4 人が pre-university education の学生, 1 人が higher professional education の学生
 - seq. 1: 2 人; seq. 2: 1 人; seq. 3: 1 人; seq. 4: 1 人

- Apparatus and materials
 - ◇ Troubleshooting task
 - 不具合のある電気回路のシミュレーションプログラム
 - それぞれの回路は複数の欠陥を持っている
 - ◇ 電池やメーターの誤接続やショート
 - ◇ 電池の容量不足や不適切な抵抗値による電流, 電圧の問題 などなど...

- タスクに課せられる条件
 - a) 外見上観察できる機能は回路が閉じているとき見ることができる
 - b) 修理された回路は少なくとも初期の回路と同じコンポーネントを含む
(回路に新しいコンポーネントを加えることは可能)
 - c) すべてのコンポーネントは正しく接続されていなければならない
 - d) スイッチが複数ある場合、全部のスイッチがオンの状態で回路が正しく機能する
- 分析に用いる回路 (Figure 1)
 - 主要な欠陥
 - ◇ 電源電圧 (3V) がランプを点灯させるのには小さすぎる
 - ◇ ランプ 1 がスイッチの状態によらずショートしている
 - 軽微な欠陥
 - ◇ メーター類が逆接続になっている (指示値が負になっている)
- ◇ Performance rating
 - 以下の観点から答案を採点 (1 つの課題についてすべてが適用されるわけではないらしい)
 - スイッチの位置と動作の関係
 - ランプの輝度
 - 最適な動作
 - ◇ Fig. 1 の回路の場合、抵抗が調光器として機能しているか
 - 各コンポーネントの接続
 - メーターの接続方向
 - 不要なコンポーネントの有無
 - 極端な数値のコンポーネントの有無
 - 配点は各項目につき 1 点、全体で 7 点満点
 - 採点形式の信頼性
 - 2 人の評価者が 15 人の被験者を採点
 - ◇ $R = 0.94$, $\rho = 0.97$
- ◇ Mental effort
 - タスク終了後に ‘very, very low effort’ ~ ‘very, very high effort’ の 9 段階で回答
(Paas, 1992; Paas, Tuovinen, Tabbers, & Van Gerven, 2003)
- ◇ Registration of eye movements
 - 眼球運動測定装置
 - ビデオベース
 - 時間解像度: 50 Hz
 - 視線解像度: 0.5° 以下
 - 刺激提示装置
 - 21 インチディスプレイ
 - 解像度: 1024 × 768 pixel
 - ◇ 刺激提示部の大きさ: 740 × 480 pixel

- 記録方法
 - ディスプレイから 70 cm の位置に被験者の頭部が来るようにヘッドレストを調整
 - ディスプレイ下部の赤外線ビデオカメラで眼球運動を測定
 - GazeTracker™ により、眼球運動およびマウス・キーボードの操作履歴を記録
- Procedure
 - ◇ 眼球運動測定装置のキャリブレーション
 - ◇ GazeTracker 記録開始
 - ◇ トラブルシューティング（‘think aloud’）
 - ◇ Mental effort の評定
- Data reduction and analysis
 - ◇ マウスクリックのデータに基づいてそれぞれのフェーズにどれだけ時間をかけたかを分類
 - ‘problem orientation’
 - スタートからスイッチが操作されるまでの間
 - ‘problem formulation and action decision’
 - 最初にスイッチが操作されてから回路の修正に着手するまで
 - ‘action evaluation and next action decision’
 - 一連の修正が完了してから次の修正に着手するまで
 - ◇ フェーズ 2, 及びフェーズ 3 におけるサブフェーズの扱い
 - 各サブフェーズを大雑把に区別するため、フェーズの所要時間を 2 等分し、前半と後半で個別に停留時間を計算する
 - ◇ 停留点
 - 停留点の特定: Dispersion-based method (Salvucci, 1999)
 - 半径 40 pixel の範囲内に 100 ms 以上停滞した場合を凝視とみなす
- Result
 - 結果について
 - ◇ サンプル数が少ないのでノンパラメトリック検定を用いた
 - ◇ Type エラーを避けるために U 検定の有意水準を 0.10 とする
 - Phase 1: ‘problem orientation’
 - ◇ 所要時間 (Table 3, Phase 1)
 - Higher > Lower (U = 4.0, p = 0.095)
 - ◇ 平均停留時間 (Table 3, Phase 1)
 - Higher < Lower (U = 4.0, p = 0.095)
 - ◇ Phase 1 の全停留時間に占めるバッテリーへの停留時間の割合 (%)
 - Higher > Lower (U = 4.0, p = 0.087)
 - Higher: 5.00-10.53-15.88

- Lower: 0.00-5.45-7.69
- ◇ スイッチとランプ 1 間を注視点が移動した回数
 - Higher = Lower (U = 5.5, p = 0.175)
 - Higher: 0.5-2.0-4.0
 - Lower: 0.0-1.0-1.0
- Phase 2: Problem formulation and action decision
 - ◇ 所要時間 (Table 3, Phase 2)
 - Higher = Lower (U = 6.0, p = 0.222)
 - ◇ 平均停留時間 (Table 3, Phase 2)
 - Higher = Lower (U = 6.0, p = 0.222)
 - ◇ Problem formulation サブフェーズにおける停留時間 (Table 3, Phase 2.1)
 - Higher > Lower (U = 4.0, p = 0.095)
 - ◇ Action decision サブフェーズにおける停留時間 (Table 3, Phase 2.2)
 - Higher = Lower (U = 12.0, p = 1.000)
- Phase 3: Action evaluation and next action decision
 - ◇ 所要時間 (Table 3, Phase 3)
 - Higher > Lower (U = 3.0, p = 0.056)
 - ◇ 平均停留時間 (Table 3, Phase 3)
 - Higher = Lower (U = 6.0, p = 0.222)
 - ◇ Action evaluation サブフェーズにおける平均停留時間 (Table 3, Phase 3.1)
 - Higher = Lower (U = 12.0, p = 1.000)
 - ◇ Action evaluation サブフェーズにおける平均停留時間 (Table 3, Phase 3.1)
 - Higher = Lower (U = 6.0, p = 0.222)
- Mean fixation duration over phase (Figure 2)
 - ◇ Lower 群
 - (サブ) フェーズ間での差はない ($\chi^2(4) = 6.40, p = 0.174$)
 - ◇ Higher 群
 - (サブ) フェーズ間での差が有意 ($\chi^2(4) = 10.40, p = 0.022$)
 - 多重比較 (Nemenyi post-hoc procedure)
 - ◇ Phase 1 < Phase 2.1, 2.2, 3.2
 - ◇ Phase 2.1 > Phase 3.1
- Combining eye movement and concurrent verbal protocol data (Table 4)
 - ◇ (録音時のエラーにより higher 群の 1 人の発話プロトコルが取得できなかった)

◇ 発話プロトコルに関する定性的な見解

● Phase 1:

➤ 所要時間の結果 (Higher > Lower) に合致する

◇ Lower 群 (Table 4: L3, L4, L5)

- 問題点を見極めることをほとんどしない
- すぐに回路の機能をテストするところに焦点が移ってしまう

◇ Higher 群 (Table 4: H1, H3)

- 回路のコンポーネントの一覧を作成する
- テストする前にその機能を予測する
 - ✓ この予測行為はバッテリーへの停留時間が長いことと関係がある？

➤ 平均停留時間の結果 (Higher < Lower) に反する

- Higher 群のほうがより完全な orientation をする傾向がある
- にもかかわらず、平均停留時間のデータは lower 群のほうがより多くの処理をしていることを示している

● Phase 2

➤ Problem formulation サブフェーズにおける平均停留時間の結果 (Higher > Lower) と関連？

◇ Higher 群はより確証をもって問題点の判断を開始するように見える

● Phase 3

➤ 所要時間の結果 (Higher > Lower) について

◇ Lower 群は彼らのとった行動の結果を higher 群ほど細かく評価していない

➤ 平均停留時間の結果 (Higher = Lower) に反する

◇ Higher 群はより完全な評価を行う傾向がある

• Discussion

➤ 所要時間 (予測 1, 2)

◇ 予測

- Higher 群は lower 群よりも problem orientation, problem formulation, action decision, evaluation に長い時間を費やす (予測 1, 及び予測 2)

◇ 結果

- ‘problem orientation’, ‘action evaluation and next action decision’ フェーズにおいて確認された
- ‘problem formulation and action decision’ フェーズにおいては確認されなかった

➤ ‘problem orientation’ フェーズでのプロトコルデータより

◇ Lower 群

- すぐに回路の機能をテストすることに焦点が移る

◇ Higher 群

- 回路の動作を予測する傾向がある

- Higher 群は回路の動作を予測することで, ‘problem formulation and action decision’ フェーズに掛かる時間を短縮している
 - ◇ 熟達度の高い人のメンタルモデルはより発達しているという仮説と一致する
- ◇ Higher 群と lower 群の差異
 - Lower 群
 - まず回路の機能をテストしてから問題を考える
 - ◇ 機能のテストは心的表象を構築するよりも認知負荷が小さい?
 - Higher 群
 - メンタルモデルの効率が良いため, 表象を構築するキャパシティーがある
 - 表象により心的に回路をテストして, 回路の動作についての予測を立てる
 - メタ認知的な違いと関連? (Schoenfeld, 1987)
 - 熟達度の低い人
 - ◇ 最初に思い浮かんだ方向性を維持
 - 熟達度の高い人
 - ◇ 最初の動作を決定するまでに比較的長い時間を費やす
 - ◇ 自分の行動によりどれだけ目標に近づいたかを評価 / モニタリング
- 平均停留時間から見た認知負荷 (予測 3, 4)
 - ◇ Higher 群と lower 群の差異
 - Higher 群
 - ‘problem orientation’ では長い時間を費やすにもかかわらず ‘problem formulation and action decision’ よりも認知負荷は低い
 - ◇ 停留時間 短い
 - 知覚的なエンコーディングのプロセス
 - ◇ 停留時間 長い
 - 問題解決プロセス
 - Lower 群
 - すべての (サブ) フェーズで認知負荷に有意差なし
 - ◇ 予想外
 - ◇ Mental effort との関連
 - Phase 1 以外では平均停留時間に差はない
 - タスク全体の Mental effort (Table 1)
 - Lower 群 > Higher 群 ($U = 0, p = 0.008$) 驚くべきこと
 - 停留時間を通して測られたものと, 総体的に感じられた心的な労力は異なるもの
- 総括
 - ◇ 被験者はランダムにシーケンスにアサインされた
 - 結果は人為的なものではない

- ◇ 眼球運動やマウス / キーボードの操作履歴を使うことの弱点
 - Phase 2 及び phase 3 のサブフェーズを弁別できない
 - 本研究では、何らかの手がかりを得るためにフェーズを 2 分割した
 - ◇ 注意深く解釈されるべき
 - 発話プロトコルデータでもサブフェーズの分類はできなかった
 - プロトコルに基づいて分割点を定義することの問題点
 - ◇ 被験者が必ずしも思い浮かんだことすべてを発話してくれるとは限らない
 - ◇ パフォーマンスの評価尺度は適切だった
 - Higher 群と lower 群を区別することが可能
 - ただし、個人が熟達化期間のどの位置にいるかを正しく評価することはできない
 - ◇ 眼球運動で測定される認知的負荷と伝統的な測定方法による認知負荷の違い
 - さらなる研究が必要
 - ◇ サンプル数を増やした検討
 - サンプル数が少ないために明確な主張ができない
- それでも...
- ◇ 発話プロトコルと眼球運動データを併用することは、暗黙的な認知プロセスについての見識を深めることを示した

Table 1. Means and standard deviations of the lower and higher expertise participants' mental effort, performance, and time-on-task scores on all experimental tasks

	Lower		Higher		Significant (two-tailed)
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Mental effort (scale 1–9)	5.28	0.42	3.63	0.54	$U = 0, p = 0.008$
Performance (scale 1–7)	4.55	0.40	5.69	0.34	$U = 0, p = 0.008$
Time-on-task (s)	359.98	45.46	163.29	55.47	$U = 0, p = 0.008$

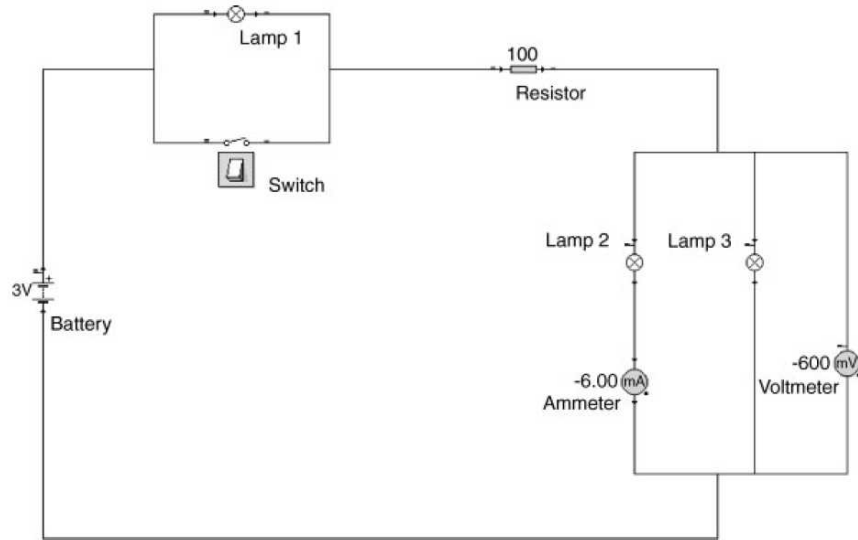


Figure 1. The troubleshooting task (text added)

Table 3. The first quartile, median (in bold), and third quartile values for the relative time spent and mean fixation duration per phase

Phase	Time (%)		Mean fixation duration (ms)	
	Lower expertise	Higher expertise	Lower expertise	Higher expertise
1	0.47– 1.03 *–5.38	4.49– 7.33 *–9.70	202.1– 217.5 *–328.6	181.2– 201.6 *–206.9
2	1.66– 2.85 –7.92	4.39– 7.34 –36.40	219.8– 245.6 –311.6	271.7– 317.7 –429.3
2.1	n.a.	n.a.	204.7– 224.3 *–250.9	255.2– 348.4 *–459.5
2.2	n.a.	n.a.	241.7– 285.0 –481.3	256.0– 330.9 –503.6
3	0.60– 3.21 *–8.76	6.46– 9.76 *–15.83	194.9– 233.9 –323.9	254.0– 291.2 –378.0
3.1	n.a.	n.a.	214.2– 252.0 –372.8	211.2– 302.6 –341.4
3.2	n.a.	n.a.	169.9– 225.3 –334.8	264.0– 327.6 –437.7

Note: Medians marked with * differ significantly between groups; n.a. = not applicable.

Table 4. Lower and higher expertise participants' verbalizations per phase

P	Phase 1	Phase 2	Phase 3
L1	Uh..well.....	Uhm...ok let's (incomprehensible)	Uhm.....
L2	Let's see first what's in the circuit... hm. Try it..	Hm, well, hm, hm yes, the ammeter is right	That lamp here..
L3	I check first whether the lamp works....	The lamp... none of the three lamps works.. first... I myself think that the power supply is too low. Check whether that's the case	So now I'm done... right? Uh..is it off now? Yes it's off right? (Experimenter: yes, the switch is open now). Hm, that's not good.... uhm... I will lower the power to 15
L4	Yes. Ok. I'll try first to see if it works. Probably not.	..Uhm..let me think Yes, I think I will here..or wait, I'll first raise the voltage, maybe there is too little	Well, that was too much... (laughs). Let's see if 12... I can always raise it later..
L5	Ok I'll see first what it does and does not do.	Ok, it does next to nothing... - 10... - I V..I was thinking first... it needs more voltage...	Nothing... again more..
H1	Ok..hm...Circuit. 3 V. So nothing will happen when I close this, voltage is too low.	Let's just raise the voltage a little	Then all three of them burn. Those glow a bit slow, uh, soft. Switch is still open... has to be closed
H2	Missing	Missing	Missing
H3	Well, what do we have here? Just two lamps with all kinds... aha, yes, so this one's wrong... oh no, this is a meter. Uh sorry, resistor. Two lamps. These are in series. In principle they would have to burn when I close this	But the voltage is too low. What do I think... what do I think now? This one has to be raised. 3.5, 4.5 should it be. But oh, there's also a resistor.	Now see what it does. That lamp can only take 9 V... and this one to. So when I'd connect them in series... this one could... 9.5... yeah... could be... they burn...
H4	Ok..uhm...first close this	It does nothing at all.	That's more like it
H5	Let's see, a switch, so we'll check first what will blow, then we can work towards that. I close it.	And nothing burns. 10 mA, that's very strange, but let's see... uh..does it say on those lamps how much... oh. 6 mA should go through it then... or am I wrong? Well, no I'm not wrong. Uh..off... I think the current is too low, so I'll raise the voltage here.	Well, those two already burn a little... this one still does nothing. Because the switch is open of course. Those lamps, how bright should they burn, because..40 mA.. 20...

Note: P = participant; L = lower expertise; H = higher expertise; m = milli; A = ampère, V = volt.

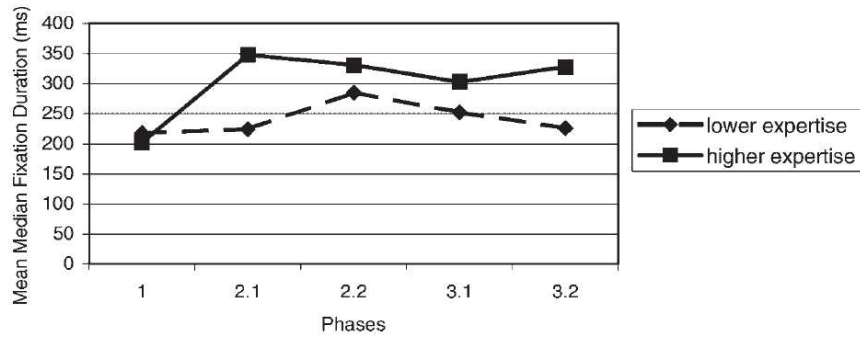


Figure 2. Fixation durations by expertise level (lower, higher) over the phases