

情報教育における作品の相互評価の評定一致度

東京女子大学 現代文化学部 廣瀬 英子
chirose@lab.twcu.ac.jp / eiko.hirose@nifty.ne.jp

1. はじめに

情報教育では他者に伝えたい情報を効果的に表現する力の育成がひとつの目的となるが、情報教育の実習のなかで作られた作品(ここではPowerPointの発表スライド)を評価する際には、単独の評定者による評価では、評定者の主観的判断が入りやすい。複数の評定者による評価を得れば、評価の客観性が高まるのではないかと考えられる。

大学の情報教育の授業において、受講生が個別に作成した作品を受講者全員に相互評価させる (peer-rating) という試みは、2001年度から続けている。本発表では、2003年度に行ったことの一部を報告する。2002年度(廣瀬, 2003)には、相互評価の際に五段階評定の各段階に割り振る作品数をあらかじめ評定者に指示した上で評価させたのに対し、2003年度は、五段階評定の各段階への作品数の配分を評定者に指示せずに評価させた。そして、評定者間での評定配分の合意の有無が評定の一致度に影響しているかどうかを検討した。また、2003年度は、総合評価点のほかに新たに8つの下位評価項目を設定して評価させた。下位項目点の評定一致度、および、下位項目点と総合評価点の関係についても考察を加えた。

2. 研究方法

2.1. スライドの作成

本研究は、都内の女子大学において、2003年11月～2004年1月にかけて、「情報科学論」の授業(2～4年生対象)のなかで行った。作品の提出と相互評価の両方を遂行した32名の受講生を分析対象者とする。受講生はPowerPointの使い方を習得したのち、作品づくりの実習を行った。

2.2. 題材

作品づくりの題材は、奥村ほか(2003)(学生の立場から遠隔授業を評価した報告: 題材 A)と、黒田ほか(2003)(大学生が親子パソコン教室を指導した報告: 題材 B)の2種類で、受講生は均等に割り振られた。ここでは、題材 A を利用して作品づくりをした受講生を A 群、題材 B を利用して作品づくりをした受講生を B 群と表記する。受講生には、「もし自分がその内容を学会で発表するとしたらどのようなスライドを用意すれば効果的であるか」を考えてスライドを作成してもらった。

2.3. 受講生による相互評価

作品は匿名化され、A 群は B 群の作った作品を、逆に B 群は A 群の作った作品を評価した。評価は8つの下位評価項目(1. 分かり易さ, 2. 要点の明確さ, 3. 論理構成, 4. アニメーション効果, 5. イラスト効果, 6. スライド枚数, 7. 指示への適格度, 8. 指示への反則度)と総合評価(全体的印象)からなり、下位評価項目は三段階評定、総合評価は五段階評定とした。

3. 結果および考察

3.1. 総合評価点の評定一致度

まず、総合評価点の内的一貫性を確かめるために、題材ごとに、16名の評価者のつけた総合評価点の α 係数を求めたところ、題材 A (B 群が評価した)で0.948、題材 B (A 群が評価した)で0.911という値が得られた。総合評価点は十分に高い信頼性が確保されたものと考えられる。

次に、作品ごとに、総合評価点の度数分布を表 1 と表 2 に示した。表中の「一致率」とは、作品ごとに最も評価の集まった評定値を選んだ評定者の割合を指す。最も高い一致率は0.75であった。それは12名の評定が一致した3作品で、表 1 の作品 d5 では1に、作品 d15 と表 2 の作品 p8 では2に集まった。表 1 と表 2 で計32作品あるうち、半数の16作品が一致率0.5以上0.75未満であった。一致率が0.5に満たないのは13作品であった。その13作品のうちの9作品は題材 A の作品であり、題材 A を評価した B 群の評定者たちは、題材 B を評価した A 群の評定者たちと比べて、ひとつの評定値にまとまらないようである。廣瀬(2003)のときは、一致率が0.5に満たない作品は21作品中4作品のみであった。廣瀬(2003)のときとの違いは、評価の際に評定値1, 2, 3, 4, 5をいくつかつけてよいかという評定配分が今回は示されず、評定者に任されたことである。今回と廣瀬(2003)とは題材、評定者が異なるため、両者を単純に比較することはできないが、配分の指示がないことによって一致率が低下することが示唆された。

もし、評定配分を正規分布にもとづいて指示していれば、1から5までの各評定段階の作品数の比率は7, 24, 38, 24, 7%となるであろうが、表 1 と表 2 の下部の比率を見ると、題材 A では5が多く、題材 B では1と2が僅かで3と5が多く、どちらも高めの評価となった。

3.2. 下位項目点の評定一緻度

下位項目点は評価のポイントが限定的である。三段階評定でもあるので、総合評価点よりも、評定者間で評価が一致しやすいことが予想される。表3には、評価項目ごとに、16作品分の一致率を要約して、平均値、標準偏差、最大値、最小値を示した。題材AとBを比較すると、題材A(B群が評価した)のほうが一致率の平均値がやや低く、標準偏差はやや大きい。下位評価項目6,7,8のような、判断に主観の入る余地の少ない項目でも一致率が劣るということは、題材Aを評価したB群の評定者たちには、ひとつの評定値にまとまりにくいという問題があるように推察される。

3.3. 下位項目点と総合評価点との関係

評定者が下位評価項目を総合して総合評価点を決めているものと仮定すれば、下位項目点を見ることによって、総合評価点の背景をいくらか理解することができるであろう。ここに、両者の関係がわかりやすく現れた例を挙げる。題材Bの作品p8は、前出の一致率が0.75の作品で、総合評価点は2に集中している。下位評価項目をみると、評価項目6,7,8に評定者全員が最高点をつけているにもかかわらず、評価項目1,4,5で12名が最低点をつけていることが響いているものと考えられる。評価項目1,4,5の要素が総合評価にもたらす影響は大きいのではないかと示唆される。

4. まとめ

本研究では、評定の際に各評定段階に割り振る作品数を指示せずに評価してもらおうと、複数の評定者の間の評定一緻率が下がる様子と、全体に高め(甘め)の評価になる様子が見られた。複数の評定者で評価するときには、評価基準のひとつとして、評定配分も含めるほうが望ましいことが、実践の中で示された。下位項目点に関しては、得られた結果をもとに、今後、下位評価項目の重みづけによる総合評価点の予測などの検討を進めていきたい。

【参考文献】

- 廣瀬英子 (2003) 情報教育における学生のパフォーマンスに対する相互評価の試み. 2003PCカンファレンス論文集, 67-68
- 黒田陽子・高須賀舞・二宮優子・奥村志沙・小田裕士・越智さやか (2003) 大学生による親子パソコン教室. 2003PCカンファレンス論文集, 99-102
- 奥村志沙・越智さやか・河上あこ・永井智恵・浜田美雪 (2003) インターネットを利用した遠隔授業のメリットと限界 学生目から見た遠隔授業. 2003PCカンファレンス論文集, 211-214

表1 総合評価点の度数分布(題材AをB群が評価したとき)

作品番号	総合評価点					横合計	一致率
	1	2	3	4	5		
D1	2	4	8	2	0	16	0.500
D2	0	1	2	6	7	16	0.438
D3	0	3	4	5	4	16	0.313
D4	0	0	4	5	7	16	0.438
D5	12	4	0	0	0	16	0.750
D6	1	9	6	0	0	16	0.563
D7	0	3	9	3	1	16	0.563
D8	1	6	8	0	1	16	0.500
D9	0	0	4	6	6	16	0.375
D10	2	6	4	3	1	16	0.375
D11	0	1	5	3	7	16	0.438
D12	0	0	4	6	6	16	0.375
D13	0	1	2	6	7	16	0.438
D14	0	1	9	2	4	16	0.563
D15	3	12	1	0	0	16	0.750
D16	1	5	7	3	0	16	0.438
縦合計	22	56	77	50	51	256	一致率
比率	8.6%	21.9%	30.1%	19.5%	19.9%	100%	平均値
							0.488

(注) 二重枠は、その作品でもっとも評価が集まった評定値を示す。

表2 総合評価点の度数分布(題材BをA群が評価したとき)

作品番号	総合評価点					横合計	一致率
	1	2	3	4	5		
P1	0	1	5	6	4	16	0.375
P2	0	1	5	8	2	16	0.500
P3	0	0	3	6	7	16	0.438
P4	0	2	11	2	1	16	0.688
P5	0	4	8	4	0	16	0.500
P6	0	0	8	4	4	16	0.500
P7	0	2	9	4	1	16	0.563
P8	2	12	2	0	0	16	0.750
P9	0	0	9	4	3	16	0.563
P10	0	0	5	3	8	16	0.500
P11	0	2	7	5	2	16	0.438
P12	2	8	5	1	0	16	0.500
P13	0	3	8	5	0	16	0.500
P14	1	3	11	0	1	16	0.688
P15	0	2	7	6	1	16	0.438
P16	0	6	9	1	0	16	0.563
縦合計	5	46	112	59	34	256	一致率
比率	2.0%	18.0%	43.8%	23.0%	13.3%	100%	平均値
							0.531

表3 下位項目点の一緻率の基礎統計量

【題材AをB群が評価】	平均値	標準偏差	最大値	最小値
評価項目1	0.625	0.129	0.875	0.438
評価項目2	0.590	0.123	0.938	0.500
評価項目3	0.578	0.077	0.750	0.500
評価項目4	0.645	0.175	1.000	0.438
評価項目5	0.559	0.111	0.750	0.375
評価項目6	0.781	0.191	1.000	0.500
評価項目7	0.840	0.097	1.000	0.688
評価項目8	0.953	0.036	1.000	0.875
【題材BをA群が評価】	平均値	標準偏差	最大値	最小値
評価項目1	0.629	0.120	0.938	0.500
評価項目2	0.582	0.094	0.813	0.438
評価項目3	0.590	0.091	0.750	0.500
評価項目4	0.660	0.155	0.938	0.438
評価項目5	0.602	0.088	0.750	0.438
評価項目6	0.992	0.022	1.000	0.938
評価項目7	0.965	0.100	1.000	0.625
評価項目8	0.996	0.016	1.000	0.938