

## BAB IV

### TEMUAN PENELITIAN DAN PEMBAHASAN

#### A. Pendahuluan

Dalam bab empat ini akan dikemukakan dua hal yaitu mengenai temuan penelitian dan pembahasan. Temuan penelitian meliputi hasil dari data utama (primer) dan data penunjang (sekunder). Data primer yang diolah yaitu skor kinerja siswa yang berasal dari penilaian guru dan penilaian *peer*. Sedangkan data penunjang (sekunder) berupa hasil angket dan wawancara tentang tanggapan siswa terhadap penerapan penilaian kinerja dengan teknik *peer assessment* pada kegiatan praktikum.

Dari data primer akan dikemukakan mengenai deskriptif statistik keseluruhan skor kinerja siswa yang dinilai oleh guru dan *peer*, analisis perbedaan skor penilaian guru dengan skor penilaian *peer*, analisis hubungan skor penilaian guru dengan skor penilaian *peer*, dan analisis *intrarater reliability* dan *interrater reliability peer assessment*. Dari data sekunder akan dikemukakan hasil angket berupa presentase jawaban siswa dan kategorinya dari setiap pertanyaan yang diberikan. Kemudian, berdasarkan temuan yang diperoleh akan dilanjutkan dengan pembahasan.

#### B. Temuan Penelitian

Penilaian kinerja siswa dilakukan berdasarkan pedoman penilaian kinerja yang telah dibuat. Penilaian kinerja dalam kegiatan pembelajaran dilakukan

berdasarkan rencana pelaksanaan pembelajaran dan skenario pembelajaran yang telah dibuat. Dalam penelitian ini yang bertindak sebagai penilai (*assessor*) adalah guru dan siswa (*peer*), dari kedua penilai ini diperoleh data berupa skor kinerja siswa yang penilaian guru dan skor penilaian *peer*. Berdasarkan skor penilaian dari lembar observasi penilaian kinerja siswa, skor ideal yang dapat dicapai oleh siswa adalah 19. Distribusi masing-masing penilaian secara lengkap ditampilkan pada lampiran D.1. Tabel 4.1 berikut menampilkan statistik deskriptif yang merupakan ringkasan dari keseluruhan skor penilaian guru dan *peer*.

**Tabel 4.1.**  
Statistik Deskriptif Skor Penilaian Guru dan Penilaian *Peer*

SERI	Statistik Deskriptif	Skor Penilaian		Bias
		Guru (G)	Peer (P)	
1	Rata-rata	14,14	16,17	G-P=2,03
	SD	3,64	2,81	
	Nilai Min	6,00	10,00	
	Nilai Max	19,00	19,00	
2	Rata-rata	14,68	16,16	G-P=1,48
	SD	3,15	2,98	
	Nilai Min	8,00	8,00	
	Nilai Max	19,00	19,00	
3	Rata-rata	15,34	15,84	G-P=0,5
	SD	2,95	3,00	
	Nilai Min	9,00	9,00	
	Nilai Max	19,00	19,00	
GABUNGAN	Rata-rata	14,72	16,06	G-P= 1,34
	SD	3,26	2,93	

Tabel 4.1 di atas menunjukkan rata-rata skor, standar deviasi, skor maksimum dan minimum penilaian guru dan *peer* dari ketiga seri. Pada seri pertama, rata-rata skor, standar deviasi, dan nilai minimum menunjukkan perbedaan yang besar. Hal ini dapat kita lihat dari selisih rata-rata skor penilaian

guru dengan *peer* adalah 2,03 serta standar deviasi dan skor minimum kedua penilaian berbeda (SD: Guru =3,64, *peer* =2,81; Skor minimum: Guru =6, *peer* =10). Pada seri kedua, rata-rata skor dan standar deviasi untuk penilaian guru dan *peer* masih menunjukkan perbedaan yang besar (G-P= 1,48), (SD: Guru =3,15, *peer* =2,98). Sedangkan, pada seri ketiga antara penilaian guru dan *peer* tidak menunjukkan perbedaan yang besar untuk skor rata-rata, standar deviasi, skor maksimum dan minimum (G-P= 0,5; SD: Guru= 2,95, *peer*= 3,00; skor minimum: Guru= 9, *peer*= 9; dan skor maksimum: Guru=19, *peer*=19). Apabila dilihat secara keseluruhan ternyata ada perbedaan yang cukup besar pada rata-rata skor penilaian *peer* dan guru (G-P = 1,34) dan standar deviasinya (SD: Guru =3,26; *peer* =2,93).

Sesuai dengan tujuan penelitian, maka skor dari penilaian guru dan *peer* yang telah diperoleh kemudian dianalisis. Adapun analisis yang dilakukan adalah analisis perbedaan rata-rata, analisis korelasi, dan analisis reliabilitas. Semua analisis dalam penelitian ini dibantu dengan menggunakan program SPSS 13 for windows.

Sebelum melakukan analisis lebih lanjut, terlebih dahulu dilakukan uji prasyarat terlebih dahulu yaitu uji normalitas. Hal ini dilakukan untuk menentukan uji hipotesis dilakukan dengan cara parametrik atau non parametrik.

Penghitungan uji normalitas dilakukan terhadap skor penilaian guru dan *peer* untuk masing-masing seri pembelajaran. Uji normalitas ini menggunakan analisis Chi kuadrat ( $\chi^2$ ). Adapun kriteria pengujiannya adalah jika  $\chi_{hitung}^2 < \chi_{tabel}^2$ , maka data terdistribusi normal. Penghitungan lengkap mengenai uji normalitas

dapat dilihat pada lampiran D.2. Rekapitulasi hasil uji normalitas disajikan pada tabel 4.2 berikut.

**Tabel 4.2** Hasil Uji Normalitas dengan Analisis Chi Kuadrat ( $\chi^2$ )

Seri	Skor Penilaian	df	$\chi^2_{hitung}$	$\chi^2_{tabel(0,05)}$	Kesimpulan
1	Guru	12	14,40	21,03	$\chi^2_{hitung} < \chi^2_{tabel}$ , terdistribusi normal
	Peer	13	19,80	22,36	$\chi^2_{hitung} < \chi^2_{tabel}$ , terdistribusi normal
2	Guru	10	14,97	18,31	$\chi^2_{hitung} < \chi^2_{tabel}$ , terdistribusi normal
	Peer	13	18,60	22,36	$\chi^2_{hitung} < \chi^2_{tabel}$ , terdistribusi normal
3	Guru	9	8,72	16,92	$\chi^2_{hitung} < \chi^2_{tabel}$ , terdistribusi normal
	Peer	16	21,83	26,30	$\chi^2_{hitung} < \chi^2_{tabel}$ , terdistribusi normal
Gabungan	Guru	22	6,40	33,92	$\chi^2_{hitung} < \chi^2_{tabel}$ , terdistribusi normal
	Peer	24	8,57	36,42	$\chi^2_{hitung} < \chi^2_{tabel}$ , terdistribusi normal

Dengan melihat hasil uji normalitas pada tabel 4.2 di atas, seluruh hasil penghitungannya menunjukkan bahwa semua skor penilaian guru dan *peer* baik tiap seri maupun gabungan terdistribusi dengan normal. Sehingga, diputuskan pengujian semua hipotesisnya menggunakan statistik parametrik.

Setelah melakukan uji normalitas, maka untuk kepentingan analisis statistik selanjutnya dilakukan uji keragaman varians (homogenitas) Adapun kriteria pengujiannya adalah jika nilai  $F_{hitung} \leq F_{tabel}$ , maka kedua data memiliki varians yang sama atau homogen. Penghitungan selengkapnya mengenai uji

homogenitas ini dapat dilihat pada lampiran D.2. Rekapitulasi hasil uji homogenitas dapat dilihat pada tabel 4.3.

**Tabel 4.3.** Hasil Pengujian Keragaman Varians (Homogenitas)

Skor Penilaian Guru dan Skor Penilaian <i>Peer</i>	$F_{hitung}$	$F_{tabel (0,1)}$	Kesimpulan
SERI 1	3,66	1,80	$F_{hitung} > F_{tabel}$ , tidak homogen
SERI 2	0,81	1,80	$F_{hitung} < F_{tabel}$ , homogen
SERI 3	0,25	1,80	$F_{hitung} < F_{tabel}$ , homogen
GABUNGAN	1,19	1,80	$F_{hitung} < F_{tabel}$ , homogen

Hasil penghitungan uji normalitas pada tabel 4.3 menunjukkan bahwa skor penilaian guru dan skor penilaian *peer* untuk seri 1 memiliki  $F_{hitung} > F_{tabel}$ , sehingga disimpulkan bahwa kedua penilaian guru dan *peer* tidak memiliki varians yang sama atau tidak homogen. Sedangkan untuk seri 2, seri 3, dan gabungan,  $F_{hitung} < F_{tabel}$ , sehingga disimpulkan bahwa kedua penilaian guru dan *peer* pada seri 2, seri 3, dan gabungan memiliki varians yang sama atau homogen.

#### 1. Perbedaan Skor Penilaian Guru dengan Skor Penilaian *Peer*

Untuk mengetahui bagaimana perbedaan antara skor penilaian guru dengan skor penilaian *peer* maka dilakukan uji t independen dua pihak. Asumsi penggunaan uji t independen adalah kedua data yang akan diuji harus berdistribusi normal dan homogen. Walaupun pada seri 1 data tidak homogen, akan tetapi dampak pelanggaran asumsi ini terhadap uji t dapat diabaikan karena jumlah subjek kedua kelompok sama besar ( $n_1=n_2$ ) (Furqon, 2004:183). Karena asumsi pengujian telah terpenuhi, maka uji t independen dapat dilakukan. Kriteria pengujian  $t_{hitung}$  adalah jika  $-t_{tabel} \leq t_{hitung} \leq t_{tabel}$ , maka tidak terdapat perbedaan yang signifikan antara skor penilaian guru dengan skor penilaian *peer*. Hasil

penghitungan dan analisis secara lengkap dapat dilihat pada lampiran D.3.

Rekapitulasi hasil uji t independen ditampilkan pada tabel 4.4 berikut ini.

**Tabel 4.4**  
Hasil Uji t Independen Skor Penilaian Guru dan Skor Penilaian *Peer*

Seri	Skor Penilaian	Rata-rata	df	$t_{hitung}$	$t_{tabel(0,05)}$	Kesimpulan
1	Guru	16,17	68	2,61	1,67	$t_{hitung} > t_{tabel}$ , ada perbedaan yang signifikan
	<i>Peer</i>	14,14				
2	Guru	16,16	68	2,01	1,67	$t_{hitung} > t_{tabel}$ , ada perbedaan yang signifikan
	<i>Peer</i>	14,69				
3	Guru	15,84	68	0,70	1,67	$-t_{tabel} \leq t_{hitung} \leq t_{tabel}$ , tidak ada perbedaan yang signifikan
	<i>Peer</i>	15,34				
Gab	Guru	14,72	68	-0,58	1,67	$-t_{tabel} \leq t_{hitung} \leq t_{tabel}$ , tidak ada perbedaan yang signifikan
	<i>Peer</i>	16,06	68			

Dengan melihat hasil uji t independen pada tabel 4.4 di atas, terlihat bahwa untuk seri pertama dan kedua  $t_{hitung} > t_{tabel(0,05)}$ , sehingga kesimpulannya adalah terdapat perbedaan yang signifikan antara skor penilaian guru dengan skor penilaian *peer* pada seri pertama dan kedua pada taraf signifikansi  $\alpha= 0,05$ . Walaupun ada perbedaan, ternyata jika diuji pada taraf signifikansi  $\alpha= 0,005$ ,  $t_{hitung} < t_{tabel(0,05)}$ . Sehingga, kesimpulannya adalah tidak terdapat perbedaan yang signifikan antara skor penilaian guru dengan skor penilaian *peer* pada seri pertama dan kedua pada taraf signifikansi  $\alpha= 0,005$ . Begitu juga untuk seri ketiga didapatkan  $t_{hitung} < t_{tabel}$ , sehingga kesimpulannya adalah tidak terdapat perbedaan yang signifikan antara skor penilaian guru dengan skor penilaian *peer* pada seri

ketiga. Untuk skor penilaian guru dan skor penilaian *peer* secara gabungan didapatkan pula untuk nilai secara keseluruhan ternyata  $-t_{tabel} \leq t_{hitung} \leq t_{tabel}$ , sehingga kesimpulannya adalah tidak terdapat perbedaan yang signifikan antara skor penilaian guru dengan skor penilaian *peer* secara gabungan.

## 2. Korelasi Skor Penilaian Guru dengan Skor Penilaian *Peer*

Untuk mengetahui ada tidaknya hubungan antara skor penilaian guru (X) dan skor penilaian *peer* (Y), maka digunakan analisis korelasi. Analisis ini akan menghasilkan koefisien korelasi yang menunjukkan pada ada atau tidak adanya hubungan, derajat hubungan dan bentuk atau arah. Teknik korelasi yang digunakan adalah teknik korelasi *pearson product moment*. Untuk menganalisis taraf signifikansinya, maka nilai  $r_{hitung}$  akan dibandingkan dengan  $r_{tabel}$  sesuai dengan taraf signifikansi yang diinginkan yaitu 5% (0,05). Kriteria pengujiannya, jika  $-r_{tabel} \leq r_{hitung} \leq r_{tabel}$ , maka tidak ada hubungan yang signifikan antara variabel skor penilaian guru (X) dan skor penilaian *peer* (Y). Penghitungan secara lengkap dapat dilihat pada lampiran D.4. Rekapitulasi hasil korelasi skor penilaian guru dengan skor penilaian *peer* ditampilkan oleh tabel 4.5 berikut ini.

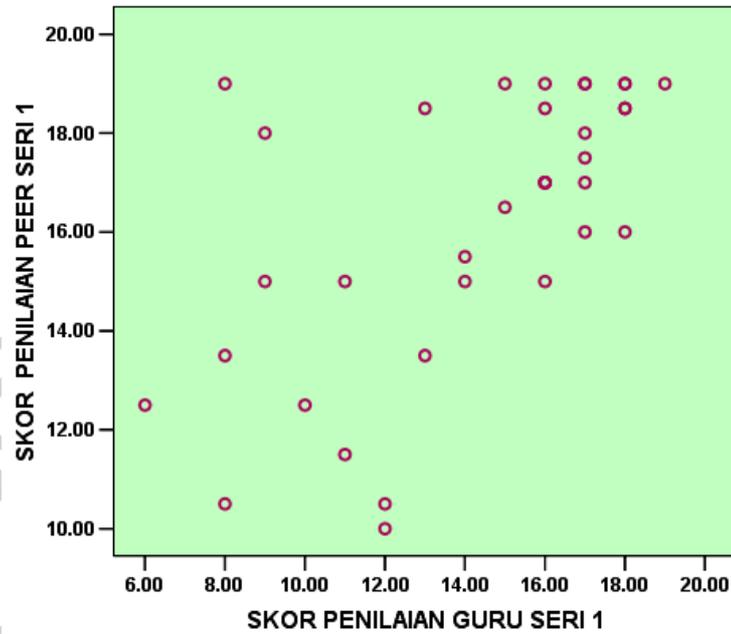
**Tabel 4.5.** Korelasi Skor Penilaian Guru dengan Skor Penilaian *Peer*

SERI	$r_{hitung}$	$r_{tabel (0,01)}$	Kesimpulan
1	0,62	0,28	$-r_{tabel} \leq r_{hitung} \leq r_{tabel}$ , ada hubungan yang signifikan
2	0,73	0,28	$-r_{tabel} \leq r_{hitung} \leq r_{tabel}$ , ada hubungan yang signifikan
3	0,92	0,28	$-r_{tabel} \leq r_{hitung} \leq r_{tabel}$ , ada hubungan yang signifikan
Gab	0,81	0,28	$-r_{tabel} \leq r_{hitung} \leq r_{tabel}$ , ada hubungan yang signifikan

Hasil penghitungan pada tabel 4.5 di atas menunjukkan bahwa untuk semua seri dan gabungan didapatkan  $-r_{tabel} \leq r_{hitung} \leq r_{tabel}$ , sehingga disimpulkan ada hubungan yang signifikan antara skor penilaian guru dengan skor penilaian *peer* untuk setiap seri. Derajat dan arah dari hubungan kedua penilaian tersebut dapat dilihat dari nilai  $r$  yang diperoleh, untuk seri pertama nilai  $r$  yang diperoleh adalah 0,62, nilai ini menunjukkan bahwa hubungan antara kedua penilaian adalah tinggi. Untuk seri kedua  $r = 0,73$  menunjukkan hubungan kedua penilaian tinggi, dan untuk seri ketiga hubungan antara kedua penilaian adalah 0,92 dengan derajat sangat tinggi. Untuk hubungan penilaian guru dan *peer* secara keseluruhan didapatkan  $r = 0,81$  dengan derajat sangat tinggi. Sedangkan untuk arah hubungan, dapat dilihat dari tanda (+) atau (-) di depan nilai  $r$ . Tanda + menunjukkan arah berbanding lurus atau korelasi positif dan tanda - menunjukkan arah berbanding terbalik atau korelasi negatif. Dari hasil penghitungan pada tabel 4.5, semua nilai  $r$  bertanda +. Sehingga dapat disimpulkan bahwa arah hubungan kedua penilaian adalah positif, searah atau sejajar. Artinya jika skor penilaian guru tinggi, maka skor penilaian *peer* tinggi, dan jika skor penilaian guru rendah, maka skor penilaian *peer* juga rendah.

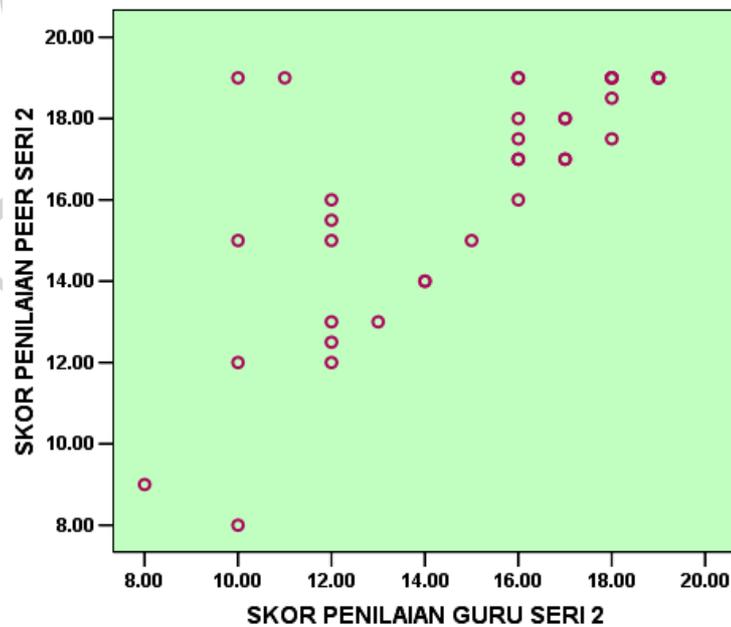
Untuk menggambarkan arah atau bentuk korelasi antara skor penilaian guru dan skor penilaian *peer*, maka digunakan diagram pencar (*scatter diagram*) dari nilai X dan Y, dimana absis X merupakan skor penilaian guru dan ordinat Y merupakan skor penilaian *peer*. Gambar 4.1, 4.2 dan 4.3 dan 4.4 menampilkan diagram pencar untuk menggambarkan korelasi skor penilaian guru (X) dan skor

penilaian *peer* (Y) untuk seri pertama, kedua, ketiga, dan keseluruhan secara berurutan.



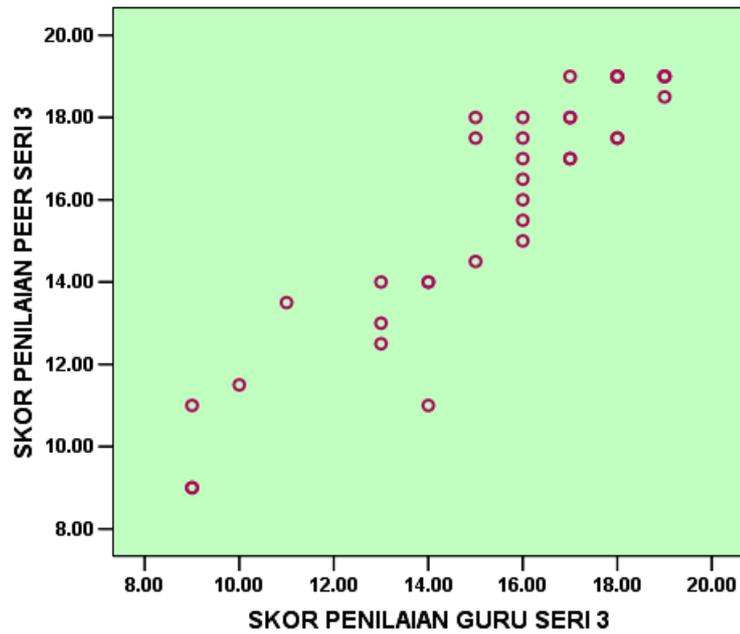
**Gambar 4.1.**

Hubungan Skor Penilaian Guru dengan Skor Penilaian *Peer* Seri 1

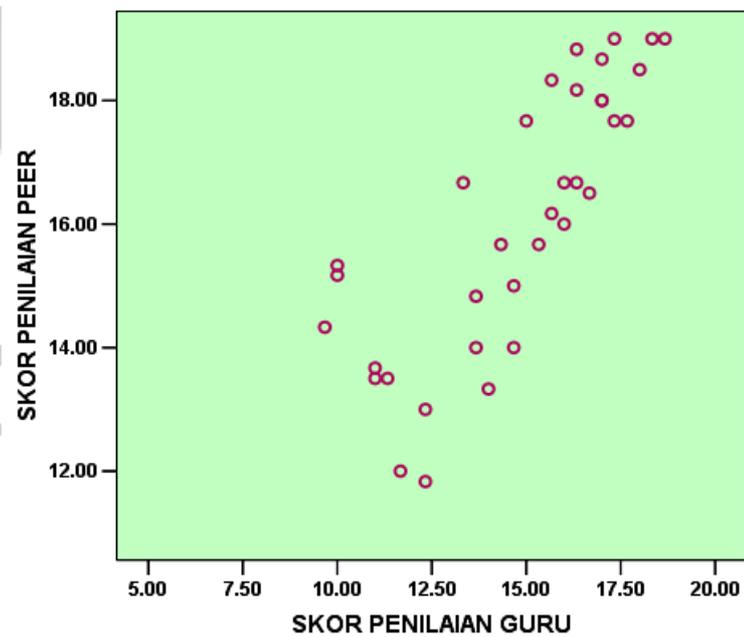


**Gambar 4.2.**

Hubungan Skor Penilaian Guru dengan Skor Penilaian *Peer* Seri 2



**Gambar 4.3.**  
Hubungan Skor Penilaian Guru dengan Skor Penilaian *Peer* Seri 3



**Gambar 4.4.**  
Hubungan Skor Penilaian Guru dengan Skor Penilaian *Peer* Secara Keseluruhan

### 3. *Interrater Reliability Peer Assessment*

Reliabilitas yang dimaksud di dalam penelitian ini adalah reliabilitas pengamat bukan reliabilitas yang dinilai atau reliabilitas tes. Reliabilitas antar penilai memberi petunjuk tentang kesepakatan dua orang penilai atau lebih dalam memberikan nilai terhadap hasil pekerjaan yang sama (Sapriati, 2005). Reliabilitas penilai (*rater*) adalah konsistensi skor yang diberikan dua orang penilai atau lebih yang independen (*interrater reliability*) dan konsistensi skor yang diberikan seorang penilai untuk waktu yang berbeda (*intrarater reliability*).

Dalam menganalisis reliabilitas *interrater reliability peer assessment* dilakukan dengan cara menghitung korelasi antar skor yang dibuat oleh dua orang *peer* dengan menggunakan rumus H.J.X Fernandes (Arikunto, 2001: 176). Penghitungan secara lengkap dapat dilihat pada lampiran D.5. Rekapitulasi hasil penghitungan *interrater reliability* disajikan pada tabel 4.6 berikut.

**Tabel 4.6.**  
Hasil Penghitungan *Interrater Reliability*

Skor Penilaian <i>Peer</i>	Koefisien kesepakatan (KK)	Interpretasi
<b>SERI 1</b>	0,77	Tinggi
<b>SERI 2</b>	0,82	Sangat tinggi
<b>SERI 3</b>	0,85	Sangat tinggi
<b>Gabungan</b>	0,81	Sangat Tinggi

Berdasarkan hasil penghitungan *interrater reliability* pada tabel 4.6 di atas, koefisien kesepakatan seri pertama adalah sebesar 0,77 dan seri kedua adalah 0,82 dan seri ketiga adalah 0,85. Sedangkan untuk koefisien kesepakatan secara

keseluruhan adalah 0,81 yang menunjukkan bahwa *interrater reliability peer assessment* adalah sangat tinggi.

#### 4. *Interrater Reliability Peer Assessment*

Sebelum melakukan analisis *intrarater reliability peer assessment*, hasil penilaian *peer* diuji normalitas dan homogenitasnya untuk menentukan teknik statistik yang akan digunakan. Hasil penghitungan dan analisis menunjukkan bahwa penilaian *peer* terdistribusi normal dan homogen, sehingga dapat menggunakan teknik analisis parametrik. Hasil penghitungan dan analisis secara lengkap terdapat pada lampiran D.6.

Selanjutnya, skor penilaian *peer* antara seri 1- seri 2, seri 2-seri 3 dan seri 1-seri 3 diuji dengan uji korelasi *pearson product moment* dan uji t berpasangan. Hasil penghitungan korelasi *pearson product moment* dan uji t berpasangan secara lengkap dapat dilihat pada lampiran D.6. Tabel 4.7 dan 4.8 berikut menampilkan hasil penghitungan uji t dan uji korelasi secara berurutan.

**Tabel 4.7.** Koefisien Korelasi Skor Penilaian *Peer*

Skor Penilaian <i>Peer</i>	N	$r_{hitung}$	$r_{tabel (0,05)}$	Keterangan
Seri 1-Seri 2	62	0,12	0,21	$-r_{tabel} \leq r_{hitung} \leq r_{tabel}$ , tidak ada hubungan yang signifikan
Seri 2-Seri 3	62	0,36	0,21	$r_{hitung} > r_{tabel}$ , ada hubungan yang signifikan
Seri 1-Seri 3	62	0,32	0,21	$r_{hitung} > r_{tabel}$ , ada hubungan yang signifikan

Kriteria pengujian koefisien korelasi adalah jika  $-r_{tabel} \leq r_{hitung} \leq r_{tabel}$ , maka tidak ada hubungan yang signifikan. Untuk seri pertama dengan kedua  $-r_{tabel} \leq r_{hitung} \leq r_{tabel}$ , sehingga dapat disimpulkan bahwa tidak terdapat hubungan

signifikan antara skor penilaian *peer* pada seri pertama dengan kedua. Untuk seri kedua dengan ketiga  $r_{hitung} > r_{tabel}$ , sehingga dapat disimpulkan terdapat hubungan yang signifikan antara skor penilaian *peer* pada seri kedua dengan ketiga. Selanjutnya, untuk seri pertama dengan ketiga  $r_{hitung} > r_{tabel}$ , sehingga dapat disimpulkan terdapat hubungan yang signifikan antara skor penilaian *peer* pada seri pertama dengan ketiga.

Untuk seri pertama dengan kedua harga koefisien pearson ( $r$ ) = 0,12, pasangan seri kedua dengan ketiga harga koefisien pearsonnya ( $r$ ) = 0,36 dan seri pertama dengan ketiga harga koefisien pearsonnya ( $r$ ) = 0,32. Ketiga koefisien ini menunjukkan bahwa derajat hubungannya rendah. Berdasarkan hasil penghitungan pada tabel 4.7, untuk semua pasangan seri didapatkan bahwa koefien korelasinya bernilai positif. Maka, dapat dikatakan bahwa arah hubungan antara skor penilaian *peer* 1 dan skor penilaian *peer* 2 untuk setiap pasangan seri adalah searah.

**Tabel 4.8.** Uji t Berpasangan Skor Penilaian *Peer*

Skor Penilaian <i>Peer</i>	df	$t_{hitung}$	$t_{tabel}$	Keterangan
Seri 1-Seri 2	61	-0,18	1,67	$-t_{tabel} \leq t_{hitung} \leq t_{tabel}$ , tidak ada perbedaan yang signifikan
Seri 2-Seri 3	61	1,42	1,67	$-t_{tabel} \leq t_{hitung} \leq t_{tabel}$ , tidak ada perbedaan yang signifikan
Seri 1-Seri 3	61	1,10	1,67	$-t_{tabel} \leq t_{hitung} \leq t_{tabel}$ , tidak ada perbedaan yang signifikan

Kriteria pengujian  $t_{hitung}$  adalah jika  $-t_{tabel} \leq t_{hitung} \leq t_{tabel}$ , maka tidak ada perbedaan yang signifikan. Tabel 4.8 di atas menunjukkan bahwa

–  $t_{\text{tabel}} \leq t_{\text{hitung}} \leq t_{\text{tabel}}$  untuk semua pasangan seri, sehingga kesimpulannya adalah tidak ada perbedaan yang signifikan penilaian *peer* untuk semua pasangan seri. Berdasarkan kedua analisis di atas maka dapat disimpulkan bahwa penilaian *peer* pada seri pertama dengan seri kedua tidak reliabel. Sedangkan untuk seri kedua dan ketiga, pertama dan ketiga, *peer assessment* reliabel.

## **5. Tanggapan Siswa terhadap Penerapan Penilaian Kinerja dengan Teknik *Peer* pada Kegiatan Praktikum**

Hasil angket yang disebar di kelas sampel penelitian mengenai penerapan penilaian kinerja dengan teknik *peer assessment* pada kegiatan praktikum mendapatkan respon yang berbeda dari siswa, hasilnya secara lengkap dapat dilihat pada lampiran C.8. Respon terhadap kegiatan pelatihan, hampir seluruhnya (97,06%) menyatakan bahwa kegiatan pelatihan sebelum kegiatan *peer assessment* mempunyai manfaat yang sangat besar. Sedangkan tanggapan siswa terhadap setuju atau tidaknya pelaksanaan penilaian kinerja dengan teknik *peer assessment* hampir seluruhnya memberikan respon yang baik, mereka setuju dengan diadakannya penilaian seperti itu (76,47%), merasa nyaman (76,47%) dan merasa tidak dirugikan dengan teknik penilaian ini (91,18%). Hal ini dikarenakan mereka mendapatkan manfaat langsung dari teknik penilaian ini. Misalnya, hampir seluruh siswa (85,29%) setuju bahwa *peer assessment* dapat membuat siswa menjadi bagian dalam penentuan nilai, hampir seluruhnya juga (94,12%) menyatakan bahwa *peer assessment* dapat membuat siswa mendapatkan kesempatan untuk memperoleh nilai tambahan selain dari tes tertulis, hampir seluruhnya (97,06 %) setuju bahwa *peer assessment* dapat melatih sikap jujur dan

objektif ketika menilai kinerja teman, hampir seluruhnya (85,29%) merasa bahwa *peer assessment* membuat siswa merasa lebih dihargai oleh teman karena bisa berkontribusi untuk penilaian kelompok, hampir seluruhnya (94,12%) dapat mengungkap kemampuan siswa yang biasanya tidak terdeteksi oleh guru, hampir seluruhnya (94,12%) merasa bahwa *peer assessment* dapat mengetahui kelebihan siswa dalam penilaian kinerja saat praktikum, seluruhnya (100%) merasa bahwa *peer assessment* dapat mengetahui kekurangan siswa ketika melakukan praktikum, hampir seluruhnya (94,12%) merasa bahwa *peer assessment* dapat membantu mengatasi rasa bosan terhadap sistem pembelajaran dan sistem penilaian yang selama ini sudah biasa dilakukan, dan hampir seluruhnya (94,12%) merasa bahwa *peer assessment* dapat membuat termotivasi untuk lebih aktif dan lebih baik lagi dalam proses belajar mengajar terutama dalam kegiatan praktikum fisika.

Selain manfaat yang diperoleh, hampir separuh siswa (32,35%) mengalami banyak kesulitan atau kendala ketika melakukan penilaian kinerja dengan menggunakan teknik *peer assessment* ini, seperti hampir seluruhnya (97,06%) merasa bahwa rasa persahabatan, rasa suka/senang, tidak senang, benci, dendam dan persaingan terhadap teman mempengaruhi dalam memberikan penilaian secara jujur/objektif, sehingga hampir separuhnya (35,29%) tidak menilai pekerjaan teman secara objektif, jujur dan apa adanya, hampir seluruhnya (82,35%) sering memberi tambahan nilai karena merasa kasihan ketika harus memberikan nilai rendah terhadap penampilan temannya, dan ada sebagian kecil (11,76%) merasa takut untuk memberikan penilaian secara objektif dan jujur terhadap kinerja yang

dilakukan temannya. Selain itu, karena rasa persahabatan pun sebagian besar (64,71%) mereka memberitahu dan membantu temannya ketika mengalami kesulitan saat melakukan praktikum.

Selain rasa persahabatan antara teman, kendala juga dirasakan karena sebagian kecil (20,59%) merasakan kehadiran observer (guru pendamping) mengganggu konsentrasi ketika praktikum. Hampir separuhnya (32,40%) siswa merasa kesulitan dalam membaca kriteria penilaian *peer assessment*. Sebagian kecil (17,60%) tidak memahami apa yang harus dinilai dari kinerja yang ditampilkan oleh temannya saat praktikum dan hampir separuhnya (47,10%) merasa bahwa konsentrasi mereka terpecah ketika melakukan penilaian, karena menunggu giliran untuk praktikum dan merasa takut tidak bisa melakukan ketika mendapat giliran untuk dinilai.

Akan tetapi, walaupun banyak kendala dalam pelaksanaannya, pendapat mereka tentang hasil penilaian kinerja adalah sebagian besar siswa (70,59%), setuju hasil *peer assessment* menjadi salah satu kriteria penentu nilai raport fisika, karena hampir seluruh siswa (85,29%) merasa puas dan merasa percaya dengan nilai yang diberikan oleh temannya.

## C. PEMBAHASAN

### 1. Perbedaan Skor Penilaian Guru dengan Skor Penilaian *Peer*

Berdasarkan konsistensi dari penilaian guru dan siswa (*peer*), siswa memberikan skor yang lebih tinggi dibandingkan skor penilaian guru ketika menilai kinerja yang dilakukan temannya. Hal ini sesuai dengan tabel 4.1 yang

menunjukkan bahwa pada setiap seri rata-rata skor penilaian *peer* selalu lebih tinggi daripada rata-rata skor penilaian guru. Berdasarkan tabel 4.1 tersebut, ditemukan bahwa perbedaan rata-rata penilaian guru dengan siswa paling besar ada pada seri pertama yaitu 2,03. Sehingga, hasil uji t menunjukkan ada perbedaan yang signifikan antara skor penilaian guru dengan skor penilaian *peer* pada seri pertama ( $t=-2,61$ ;  $\alpha= 0,05$ ). Begitu pula pada seri kedua, perbedaan rata-rata kedua penilaian masih besar yaitu 1,48 dan berdasarkan uji t masih terdapat perbedaan yang signifikan ( $t=2,01$ ;  $\alpha= 0,05$ ). Akan tetapi, untuk seri ketiga perbedaan rata-rata kedua penilaian tidak terlalu besar yaitu 0,5 dan berdasarkan uji t, memang tidak terdapat perbedaan yang signifikan antara kedua penilaian tersebut ( $t= 0,70$ ;  $\alpha= 0,05$ ). Selanjutnya, apabila dilihat secara keseluruhan ternyata hasilnya menunjukkan bahwa tidak terdapat perbedaan yang signifikan antara skor penilaian guru dengan skor penilaian *peer* ( $t= -0,58$ ,  $\alpha=0,05$ )

Ada kesesuaian hasil penelitian ini dengan penelitian yang dilakukan oleh Kilic dan Cakan (2006), ditemukan bahwa rata-rata penilaian siswa lebih tinggi dibandingkan penilaian guru. Keadaan ini dinamakan '*over significant outliers*' yang merupakan suatu keadaan dimana *peer* memberikan penilaian sesuai dengan keinginannya sendiri yang mengakibatkan perbedaan signifikan dari rata-rata penilaian yang diberikan oleh kelompok lain atau dari penilaian yang diberikan oleh guru (Lie dan Angelique, 2003).

Faktor yang menyebabkan terjadinya *over significant outliers* pada seri pertama adalah:

- a. Siswa merasa belum terbiasa melakukan penilaian dan merasa gugup karena pada saat latihan hasil penilaiannya tidak diperhitungkan. Sedangkan, untuk seri pertama dan seterusnya hasil penilaiannya diperhitungkan.
- b. Metode baku guna menghitung tingkah laku belum dilakukan dengan baik, sehingga terjadi kesalahpahaman antara observer dengan *peer*. Misalkan ada siswa yang mengulang-ngulang proses praktikum, siswa tersebut pada awal praktikum merangkai alat dengan salah dan *peer* menilainya. Beberapa waktu kemudian, dia mengubah rangkaianannya sehingga menjadi benar dan observer menilainya. Hal ini menyebabkan perbedaan penilaian antara *peer* dengan observer. Solusinya adalah dengan membagi waktu pengamatan menjadi beberapa bagian waktu yang singkat untuk tiap indikator. Peristiwa ini tidak diduga sebelumnya, karena pada saat proses pelatihan tidak ada masalah dari siswa. Hal ini terjadi karena siswa pada saat pelatihan tidak didampingi oleh observer.
- c. Siswa merasa gugup dengan kehadiran observer.
- d. Siswa sendiri tidak memahami apa yang sedang dilakukan oleh temannya saat praktikum, khususnya terjadi saat merangkai alat. Jadi, ternyata siswa belum sepenuhnya bisa membaca gambar rangkaian percobaan pada LKS dan membandingkan dengan rangkaian sebenarnya, sehingga membuat siswa merasa kesulitan membuat rangkaian listrik dan melakukan penilaian.
- e. Subjektifitas dan adanya persekongkolan antar siswa yang tidak dapat dihindari. Hal ini terjadi karena mereka ingin mendapatkan nilai yang baik.

- f. Konsentrasi mereka terpecah ketika melakukan penilaian, karena menunggu giliran untuk praktikum dan merasa takut tidak bisa mengerjakan ketika mendapat giliran untuk dinilai.

Faktor-faktor yang terjadi di atas mengakibatkan *generosity error*, yaitu kecenderungan untuk memberikan keuntungan kepada subyek apabila penilai merasa ragu-ragu (Furchan, 2004: 276). Sehingga, nilai siswa dari *peer* cenderung selalu lebih besar daripada nilai dari observer atau guru. Untuk menghindari hal tersebut terulang lagi pada seri kedua, maka guru memberikan penjelasan kembali dan diskusi tentang prosedur penilaian, tujuan dari penilaian *peer assessment* dan hasil penilaian. Dengan adanya diskusi tentang hasil penilaian, sedikitnya memberikan pengaruh terhadap penilaian *peer*. Hal ini dapat dilihat pada seri kedua, walaupun faktor-faktor yang timbul pada seri pertama masih ada, akan tetapi ada perubahan perbedaan rata-rata penilaian guru dengan *peer* yaitu semakin berkurang jika dibandingkan dengan seri pertama.

Ketika proses diskusi, guru harus membahas alasan siswa memberikan jawaban “ya” atau “tidak” untuk temannya. Diskusi seperti ini merupakan kesempatan guru untuk memberikan umpan balik dan mengajak siswa untuk berfikir lebih kritis. Selain itu, kehadiran guru dalam penilaian ini sebagai pihak penengah menanamkan rasa tanggung jawab ke dalam diri siswa ketika melaksanakan proses penilaian, karena guru hadir untuk memastikan kewajaran dari penilaian yang diberikan oleh *peer*. Akan tetapi, hal yang paling penting dari adanya pembahasan tentang penghitungan pemberian nilai, siswa dilibatkan dalam proses berfikir kritis dan belajar untuk mengemukakan pendapat.

Dengan terus-menerus diadakan diskusi tentang hasil penilaian, maka siswa akan tetap berada pada prosedur yang benar dan meningkatkan kepercayaan diri mereka untuk terlibat dalam proses penilaian. Sehingga siswa akan melakukannya dengan lebih serius. Hal ini dapat dilihat pada seri ketiga, walaupun masih ada penilaian yang subjektif, tapi dalam taraf nyata 5% tidak ada perbedaan yang signifikan antara penilaian guru dan *peer*. Artinya, siswa sudah mulai menilai secara jujur dan objektif.

## 2. Korelasi Skor Penilaian Guru dengan Skor Penilaian *Peer*

Hasil analisis korelasi *pearson product moment* menunjukkan bahwa besar koefisien korelasinya semakin meningkat ( $r_1 = 0,62$ ;  $r_2 = 0,73$ ;  $r_3 = 0,92$ ). Ketinggian hubungan antara penilaian guru dengan penilaian *peer* dapat dilihat dari diagram pencar pada gambar 4.1, 4.2 dan 4.3. Pada gambar 4.1, dapat terlihat bahwa titik-titik dalam diagram pencarnya menyebar luas, artinya hubungan antara penilaian guru dengan *peer* relatif lemah. Pada gambar 4.2 dan 4.3 dapat terlihat bahwa titik-titik dalam diagram pencarnya mulai membentuk pita sempit, artinya hubungan antara penilaian guru dengan *peer* semakin tinggi. Dari gabungan ketiga seri, didapatkan  $r = 0,80$  dengan hubungan yang sangat tinggi. Hal ini dapat dilihat juga pada gambar 4.4, titik-titik dalam diagram pencarnya mulai membentuk pita sempit, sehingga jika ditarik garis lurus melalui pita tersebut titik-titik tersebut akan berada di dekat garis itu.

Lemahnya hubungan penilaian guru dengan siswa pada seri 1 dan 2 yang mengakibatkan adanya perbedaan disebabkan karena masih banyak *peer* yang memberikan penilaian secara berlebihan atau '*overmarking*'.

Pond, et al. (Shimura, 2006: 105) menyatakan bahwa terjadinya *overmarking* oleh *peer* sebagai nilai karena unsur pertemanan atau “*friendship marking*”. Falchikov (Shimura, 2006: 105) menyatakan dengan tegas bahwa nilai karena pertemanan ini disebabkan karena siswa merasa sulit untuk mengkritisi temannya. Pernyataan ini sesuai dengan pembahasan sebelumnya tentang faktor-faktor yang menyebabkan perbedaan penilaian *peer* dengan guru. Pada mulanya, penilaian siswa masih bersifat subjektif akan tetapi semakin lama akan timbul kesadaran dari dalam diri siswa. Kenyataan ini dapat dilihat dari hubungan penilaian guru dengan *peer* semakin meningkat dan tidak ada perbedaan secara signifikan pada seri ketiga.

Isaacs (1999) menyatakan bahwa terjadinya *overmarking* timbul karena siswa merasa kurang percaya diri ketika memberikan nilai dan kurang berpengalaman untuk melakukan penilaian. Atau menurut Furchan (2004: 276) hal tersebut juga mungkin disebabkan oleh beberapa kesalahan yang paling sering terjadi dalam penilaian yaitu:

- a. Pengaruh lingkaran (*halo effect*), yaitu kesalahan sistematis yang terjadi apabila dalam menilai aspek khusus tingkah laku, penilai terpengaruh oleh kesan umum subyek yang bersangkutan. Kesan umum ini dibawa dari satu pertanyaan dalam skala ke pertanyaan berikutnya. Sebagai contoh, seorang guru mungkin menilai bahwa murid yang prestasi akademiknya baik akan unggul pula dalam hal kecerdasan, popularitas, kejujuran, ketekunan, dan semua aspek kepribadian lainnya;

- b. Terlalu murah hati (*generosity error*), yaitu kecenderungan untuk memberikan keuntungan kepada subyek apabila penilai merasa ragu-ragu. Maskudnya, apabila peneliti merasa tidak yakin, ia cenderung memberikan nilai yang menguntungkan orang yang dinilai itu.
- c. Terlalu kikir (*errorr of severity*), yaitu kecenderungan untuk menilai semua subyek terlalu rendah pada semua ciri.
- d. Bersikap netral (*errorr of central tendency*), yaitu kecenderungan untuk menghindari penilaian yang ekstrim dan menilai semua individu dengan skala tengah-tengah.

Sekali lagi, dengan adanya penjelasan kepada siswa bahwa yang menjadi fokus kegiatan *peer assessment* adalah proses penilaian bukan kepada hasil penilaian sangatlah penting. Guru harus menjelaskan bagaimana dasar pemikiran dan manfaat-manfaat yang dapat diperoleh dari *peer assessment* kepada siswa dan dari sisi tujuan formatifnya (Lie dan Angelique, 2003), yang sebenarnya bahwa tujuan utamanya adalah untuk mendapatkan *feedback* secara langsung. Dengan pengertian-pengertian seperti itu, maka siswa akan terlibat lebih aktif dan melakukannya dengan lebih serius. Hal ini didukung oleh fakta bahwa hampir seluruh siswa (94,12%) merasa bahwa *peer assessment* dapat membuat termotivasi untuk lebih aktif dan lebih baik lagi dalam proses belajar mengajar terutama dalam kegiatan praktikum fisika.

### **3. Interrater Reliability Peer Assessment**

Hasil penghitungan pada tabel 4.6 menunjukkan bahwa semakin lama koefisien kesepakatannya atau indeks reliabilitas antar penilai semakin meningkat

dari 0,77 pada seri pertama menjadi 0,82 pada seri kedua dan menjadi 0,85 pada seri ketiga. Hal ini menunjukkan kedua *peer* semakin lama mempunyai persepsi yang sama tentang apa yang sedang mereka nilai dari pekerjaan temannya.

Kriteria sangat berpengaruh terhadap *interrater reliability*, karena tanpa adanya kriteria yang memandu proses penilaian, penilai tidak memiliki persepsi yang sama. Dengan memperjelas kriteria kinerja dan materi lain yang tercantum pada pedoman penilaian serta melakukan penilaian oleh dua orang atau lebih (untuk tes kinerja atau perilaku) untuk menjamin konsistensi penskoran (Moskal dan Leydens, 2006:6) akan meningkatkan *interrater reliability*. Sehingga, kegiatan pelatihan dan pemberian penjelasan secara bertahap tentang prosedur penilaian harus dilakukan. Akan tetapi selain itu, suasana hati penilai, persepsi penilaian, dan situasi saat penilaian juga berpengaruh terhadap proses penilaian.

Dari hasil-hasil temuan di atas dapat diambil kesimpulan bahwa pada taraf awal seri pertama dan kedua, hasil penilaian yang dilakukan oleh siswa tidak dapat langsung dipercayai dan digunakan, karena dua alasan utama. Pertama, karena siswa belum terbiasa dan terlatih, sehingga sangat terbuka kemungkinan bahwa siswa banyak melakukan kesalahan dalam penilaian. Kedua, faktor subjektifitas siswa yang sulit dihindari, karena terdorong oleh rasa persahabatan atau permusuhan dan adanya persengkongkolan untuk mendapatkan nilai yang baik.

Akan tetapi semakin lama, hasil penilaian siswa dapat dipercaya. Hal ini dapat dilihat pada seri ketiga bahwa hasilnya lebih baik. Siswa telah terlatih dalam melakukan penilaian dengan teknik *peer assessment* secara baik, objektif, dan

jujur, sehingga apabila hal ini terjadi maka akan sangat membantu meringankan beban tugas guru. Hasil penilaiannya pun dapat dijadikan dasar bagi guru untuk memberikan nilai kompetensi siswa.

#### **4. *Intrarater Reliability Peer Assessment***

Hasil *intra-rater reliability peer assessment* atau konsistensi skor penilaian *peer* yang diberikan oleh untuk waktu yang berbeda menunjukkan bahwa untuk seri pertama dengan seri kedua tidak ada perbedaan yang signifikan ( $t = -0,18$ ;  $\alpha = 0,05$ ), dengan koefisien korelasinya menunjukkan bahwa hubungannya tidak signifikan ( $r = 0,12$ ;  $\alpha = 0,05$ ) dengan derajat hubungan sangat rendah atau dianggap tidak ada hubungan sama sekali. Hal ini menunjukkan bahwa penilaian *peer* pada seri pertama dengan seri kedua tidak reliabel. Sedangkan untuk seri kedua dengan seri ketiga, hasil uji t berpasangan menunjukkan bahwa tidak ada perbedaan yang signifikan antara skor penilaian *peer* seri kedua dengan seri ketiga ( $t=1,42$ ;  $\alpha= 0,05$ ) dan koefien korelasinya menunjukkan bahwa ada hubungannya yang signifikan ( $r = 0,36$ ;  $\alpha= 0,05$ ) dengan derajat hubungan yang rendah, karena tidak ada perbedaan yang signifikan dan hubungannya signifikan, maka penilaian *peer* pada seri kedua dengan seri ketiga reliabel. Penghitungan yang sama dilakukan untuk seri pertama dengan ketiga, yang hasilnya menunjukkan tidak ada perbedaan yang signifikan antara skor penilaian *peer* seri pertama dengan seri ketiga ( $t =1,10$ ;  $\alpha= 0,05$ ) dan koefien korelasinya menunjukkan bahwa ada hubungannya yang signifikan ( $r= 0,32$ ;  $\alpha= 0,05$ ) dengan derajat hubungan rendah, yang berarti skor *peer* pada seri pertama dengan seri ketiga juga reliabel.

Seri pertama dan kedua tidak reliabel berarti kedudukan relatif tiap-tiap skor penilaian *peer* pada seri kedua tidak mendekati kedudukannya dalam tes yang pertama. Hal ini dapat disebabkan oleh kondisi *peer* atau kondisi cara penilaian. Misalnya, dari pembahasan sebelumnya kita dapat mengetahui bahwa pada seri pertama ada kesalahpahaman prosedur penilaian dan faktor-faktor yang menyebabkan perbedaan penilaian *peer* dari penilaian observer. Akan tetapi, dengan adanya diskusi antara guru dengan *peer*, pada seri kedua keadaannya menjadi berubah. Hal ini yang menyebabkan penilaian *peer* menjadi tidak ajeg atau tidak reliabel.

Penilaian *peer* pada seri ketiga dengan kedua dan pertama dengan kedua, memiliki reliabilitas rendah. Secara teori, ada tiga sumber penyebab rendahnya reliabilitas, yaitu dari faktor instrumen, siswa dan penilai. (Rudner dan Schafer dalam Sapriati, 2005). Faktor instrumen mencakup definisi perilaku, prosedur pengumpulan data, penentuan waktu dan tempat pengamatan, cara kuantifikasi perilaku, dan skala skor. Faktor siswa antara lain terdiri atas perilaku siswa yang dipengaruhi tujuan pendidikan, pengalaman atau kebiasaan asesmen dan terbiasa tidaknya dengan penilai asing. Sedangkan faktor penilai antara lain adalah kurang terbiasa dengan sistem pengamatan, waktu pengamatan tidak cukup, kurang konsentrasi, terjadi perubahan kriteria selama penilaian, harapan yang besar tentang perilaku yang diamati, motivasi penilaian selama menilai dan penilai dipengaruhi oleh bias.