

## **RELIBILITAS INTER RATER INSTRUMENT TES MATEMATIKA MENGUNAKAN RATER**

**Destiniar**

Universitas PGRI Palembang  
Email: [destiniarpgri@yahoo.co.id](mailto:destiniarpgri@yahoo.co.id)

### **Abstract**

*The purpose of this study is to know whether there are differences in reliability coefficients inter rater mathematics tests assessed by teachers of mathematics SMP and lecturer of mathematics education. Rater used in this study consisted of 20 teachers of mathematics junior and 20 lecturers of mathematics education while the instrument is a class IX mathematical test instrument in the form of multiple choices. The research method used is experiment and data analysis is done by using T test statistic. The result obtained there is no difference of inter rater reliability coefficient about mathematics test which is assessed by mathematics teacher of SMP and lecturer of mathematics education.*

**Keywords:** Reliability, Inter Rater, Test

### **Abstrak**

*Tujuan penelitian ini adalah untuk mengetahui adakah perbedaan koefisien reliabilitas inter rater soal tes matematika yang dinilai oleh guru matematika SMP dan dosen pendidikan matematika. Rater yang digunakan pada penelitian ini terdiri dari 20 orang guru matematika SMP dan 20 orang dosen pendidikan matematika sedangkan instrument yang dinilai adalah instrument tes matematika kelas IX yang berbentuk pilihan ganda. Metode penelitian yang digunakan adalah eksperimen dan analisis data dilakukan dengan menggunakan statistic uji t. Hasil yang diperoleh tidak ada perbedaan koefisien reliabilitas inter rater soal tes matematika yang dinilai oleh guru matematika SMP dan dosen pendidikan matematika.*

**Kata kunci :** Reliabilitas, inter rate, tes

## **1. PENDAHULUAN**

Pada bidang pendidikan pengukuran merupakan hal sangat penting, dan terdapat sangat banyak objek yang dapat diukur diantaranya hasil belajar siswa. Untuk melakukan pengukuran terhadap hasil belajar maka diperlukanyang namanya instrumen hasil belajar. Menurut Kartowagiran (2011: 2) bahwa sebelum instrumen tes dibuat, terlebih dahulu harus ditetapkan tujuan tes tersebut, membuat kisi-kisi tes kemudian menyusun atau merakit soal tes. Jenis tes yang paling sering dan umum digunakan adalah tes pilihan ganda. Hal ini dikarenakan tes pilihan ganda mudah untuk mengukur beberapa aspek di dalam pendidikan, seperti pengetahuan, pemahaman. Tes pilihan ganda ini biasanya digunakan untuk mengukur suatu kemampuan tertentu yang

memiliki cakupan materi yang sangat banyak dan luas, Instrumen tes hasil belajar dianggap baik bila memenuhi kriteria valid, reliabel dan *usable* (Kusaeri, 2014: 50). Sebuah instrumen tes dikatakan valid apabila tes tersebut mengukur dengan tepat apa yang hendak diukur (Widoyoko, 2015: 129). Reliabilitas alat ukur adalah ketepatan atau keajegan alat ukur dalam mengukur apa yang diukurinya. Suatu tes dikatakan reliabel jika digunakan secara berulang dengan kondisi yang relatif sama, maka hasil yang diperoleh juga akan tetap sama atau konsisten (Djanuarsih, 2012). *usable* memiliki makna bahwa penilaian yang digunakan praktis prosedurnya. (Kusaeri, 2014: 51)

Berdasarkan hasil wawancara terbatas dengan beberapa guru diperoleh informasi bahwa instrument tes yang mereka susun atau kembang tidak melalui proses validasi tetapi langsung mereka gunakan (12 Januari 2016). Jika instrument tes yang digunakan tidak dipersiapkan dengan baik maka tentunya hasil pengukuran yang diperoleh juga kurang baik. Sehingga instrument tes yang langsung digunakan ini juga perlu diteliti agar instrument tersebut dapat mengukur dengan tepat apa yang hendak diukur.

Salah satu cara yang dapat dilakukan untuk melakukan proses validasi instrumen ini yaitu dengan menggunakan *rater*. Menurut Widhiarso (2010) menggunakan *rater* dapat meningkatkan kualitas instrument. Menurut Aisyah (2016) dalam menentukan siapa yang menjadi *rater* sangat perlu diperhatikan bidang keahlian atau latar belakang *rater* haruslah relevan dengan instrument yang akan dinilai.

Pada penelitian ini yang akan dinilai adalah instrument tes matematika siswa kelas IX, sehingga disini muncul pertanyaan adakah perbedaan koefisien reliabilitas *inter rater* soal tes matematika yang dinilai oleh guru matematika SMP dan dosen pendidikan matematika?

## 2. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan metode komparatif. Penelitian ini dilakukan untuk mengetahui perbedaan koefisien reliabilitas *inter rater* antara *rater* yang berasal dari guru matematika SMP dan dosen Pendidikan Matematika. Penelitian ini dilakukan diberbagai tempat di kota Palembang, hal ini disebabkan yang menjadi *rater* pada penelitian sebanyak 20 orang guru matematika SMP yang berasal dari berbagai SMP dan 20 orang dosen pendidikan matematika yang

berasal dari dosen FKIP Universitas PGRI Palembang dan dosen FKIP Universitas Sriwijaya Palembang.

Semua *rater* yang berjumlah 40 orang diminta untuk memberikan penilaian pada instrument tes matematika kelas IX berbentuk pilhan ganda sebanyak 40 soal dengan 4 pilihan jawaban. Setelah itu hasil penilaian dikelompokkan menjadi 2 kelompok yaitu 20 penilaian dari kelompok guru matematika SMP dan 20 penilaian dari kelompok dosen pendidikan matematika. Selanjutnya dari masing-masing kelompok diambil secara acak 6 penilaian yang kemudian dihitung koefisien reliabilitas *inter rater* dengan menggunakan metode Fleiss Kappa sehingga didapat satu nilai koefisien reliabilitas *inter rater* dari masing-masing kelompok. Hal ini dilakukan pengulangan sebanyak 20 kali sehingga dari masing-masing kelompok diperoleh 20 nilai koefisien reliabilitas *inter rater*.

Setelah data terkumpul maka dilanjutkan dengan analisis data. Sebelum data dianalisis dilakukan uji prasyarat yaitu uji normalitas dan homogenitas. Menurut Kadir (2015) setelah dilakukan uji prasyarat barulah data dianalisis dengan menggunakan uji perbedaan dua rata-rata. Uji Normalitas dan homogenitas ini dilakukan dengan berbantuan program SPSS versi 20. Setelah itu untuk uji perbedaan dilakukan uji T juga berbantuan program SPSS versi 20.

### 3. HASIL DAN PEMBAHASAN

#### Hasil Penelitian

Hasil yang diperoleh pada penelitian ini berupa dua kelompok data yaitu data dari guru Matematika SMP dan dosen Pendidikan Matematika. Berikut ini data disajikan dalam bentuk tabel 1.

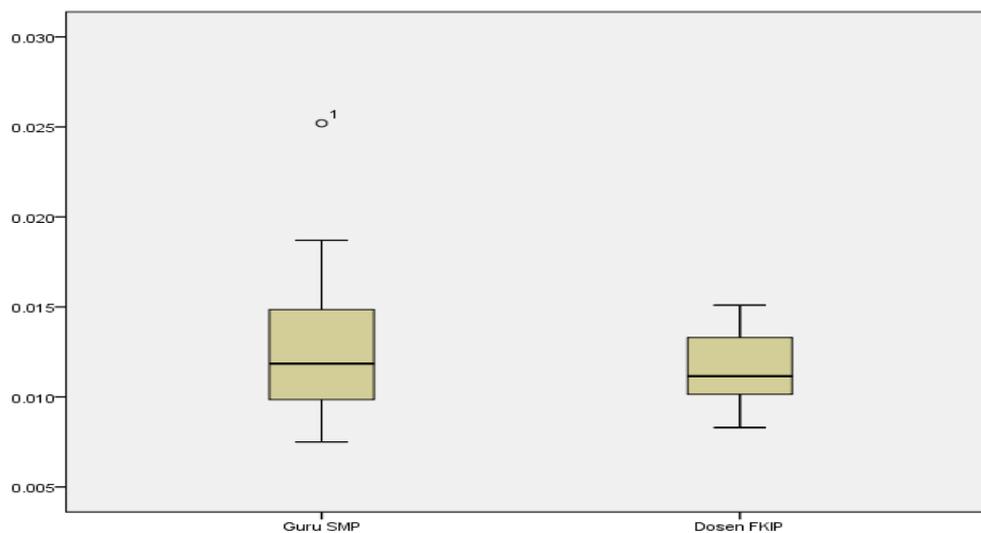
**Tabel 1 Data koefisien reliabilitas interrater**

GURU	DOSEN
0.0252	0.0151
0.0134	0.0110
0.0094	0.0106
0.0075	0.0110
0.0090	0.0140
0.0123	0.0087
0.0154	0.0083
0.0120	0.0147

0.0113	0.0117
0.0153	0.0122
0.0144	0.0135
0.0140	0.0113
0.0161	0.0096
0.0090	0.0131
0.0103	0.0107
0.0117	0.0135
0.0105	0.0108
0.0094	0.0097
0.0110	0.0122
0.0187	0.0089

Data koefisien reliabilitas *inter rater* yang bersal dari guru berjumlah 0,2558 dengan nilai rata-rata koefisien reliabilitas *inter rater* sebesar 0,0128 koefisien reliabilitas *inter rater* yang tertinggi sebesar 0.0252 dan yang terendah adalah 0,0075 sehingga diperoleh rentangnya sebesar 0,0177.

Untuk data koefisien reliabilitas *inter rater* yang berasal dari dosen pendidikan matematika berjumlah 0,2306 dengan nilai rata-rata koefisien reliabilitas *inter rater* sebesar 0,0115 koefisien reliabilitas *inter rater* yang tertinggi sebesar 0,0151 dan yang terendah adalah 0,0083 sehingga diperoleh rentangnya sebesar 00,68. Untuk melihat sebaran data dari kedua kelompok tersebut berikut disajikan dalam bentuk boxplot.



**Gambar.1 data koefisien reliabilitas tahap awal**

Dari gambar 1 dapat terlihat bahwa ada dua kelompok data yaitu data dari kelompok guru matematika SMP dan kelompok dosen pendidikan matematika. Pada gambar *boxplot* ini terdapat beberapa persamaan dan perbedaan. Panjang *box* atau kotak pada kedua kelompok data ini berbeda, *box* untuk data yang berasal dari guru matematika lebih panjang dari pada *box* data yang berasal dari dosen pendidikan matematika. Semakin panjang *box* ini menunjukkan bahwa data semakin menyebar. Median pada kedua kelompok data ini juga berbeda, hal ini terlihat dari garis yang berada pada *box*. Pada *box* data yang berasal dari guru kelihatan bahwa data tidak simetris, hal ini terlihat dari salah satu dari *whisker* lebih panjang. Pada data yang berasal dari guru matematika terdapat satu data yakni data ke 1 yang berada dibagian atas *whisker* hal ini menunjukkan bahwa nomor 1 ini merupakan data yang *outlier*.

Menurut Filzomer (2005) data outlier dapat memungkinkan terjadinya bias pada parameter yang diestimasi, dengan menghilangkan data outlier tidak akan menghilangkan informasi yang akan diukur. Mengacu pada pendapat Peter Filzomer di atas maka data yang outlier dikeluarkan dari analisis sehingga data yang berasal dari guru matematika tinggal 19 data sedangkan data yang berasal dari dosen pendidikan matematika berada sebanyak 20 data.

Berikut ini data disajikan dalam bentuk tabel :

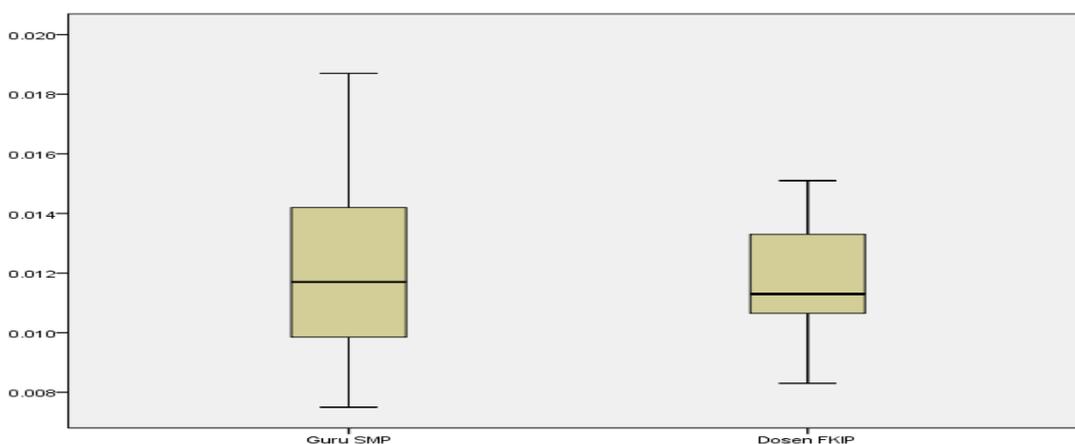
**Tabel 2 Data koefisien reliabilitas interrater**

GURU	DOSEN
0.0134	0.0151
0.0094	0.0110
0.0075	0.0106
0.0090	0.0110
0.0123	0.0140
0.0154	0.0087
0.0120	0.0083
0.0113	0.0147
0.0153	0.0117
0.0144	0.0122
0.0140	0.0135
0.0161	0.0113
0.0090	0.0096
0.0103	0.0131
0.0117	0.0107
0.0105	0.0135

0.0094	0.0108
0.0110	0.0097
0.0187	0.0122
	0.0089

Data koefisien reliabilitas *inter rater* yang bersal dari guru berjumlah 0,2306 dengan nilai rata-rata koefisien reliabilitas *inter rater* sebesar 0,0121 koefisien reliabilitas *inter rater* yang tertinggi sebesar 0.0187 dan yang terendah adalah 0,0075 sehingga diperoleh rentangnya sebesar 0,0112.

Untuk data koefisien reliabilitas *inter rater* yang berasal dari dosen pendidikan matematika berjumlah 0,2306 dengan nilai rata-rata koefisien reliabilitas *inter rater* sebesar 0,0115 koefisien reliabilitas *inter rater* yang tertinggi sebesar 0,0151 dan yang terendah adalah 0,0083 sehingga diperoleh rentangnya sebesar 00,68. Untuk melihat sebaran data dari kedua kelompok tersebut berikut disajikan dalam bentuk *boxplot*.



**Gambar.2 data koefisien reliabilitas tahap akhir**

Dari gambar 2 dapat dilihat bahwa *box* data dari guru matematika lebih besar daripada *box* data yang berasal dari dosen pendidikan matematika, ini menunjukkan bahwa data yang berasal dari guru lebih menyebar dibandingkan dengan data yang berasal dari dosen pendidikan matematika. Pada *boxplot* ini juga terlihat bahwa tidak ada lagi data yang *outlier* sehingga data inilah yang akan dianalisis.

Sebelum data dianalisis dilakukan terlebih dahulu uji prasyarat yaitu uji normalitas dan uji homogenitas data. Uji normalitas data dan homogenitas data

dilakukan dengan menggunakan bantuan *software* SPSS versi 20. Hasil uji normalitas dapat dilihat pada tabel berikut:

**Tabel 3. Uji Normalitas data**

One-Sample Kolmogorov-Smirnov Test			
		Guru SMP	Dosen FKIP
N		19	20
Normal Parameters <sup>a,b</sup>	Mean	.012142	.011530
	Std. Deviation	.0029256	.0019892
Most Extreme Differences	Absolute	.110	.105
	Positive	.110	.105
	Negative	-.089	-.089
Kolmogorov-Smirnov Z		.480	.470
Asymp. Sig. (2-tailed)		.975	.980

a. Test distribution is Normal.  
 b. Calculated from data.

Dari tabel 3 dapat dilihat bahwa untuk guru matematika diperoleh Sig. sebesar  $0,975 > 0,05$  dan untuk dosen pendidikan matematika diperoleh sig. sebesar  $0,980 > 0,05$ . Ini menunjukkan bahwa data berdistribusi normal. Untuk uji homogenitas dapat dilihat pada tabel berikut:

**Tabel 4. Uji homogenitas data**

Levene's Test of Equality of Error Variances <sup>a</sup>			
Dependent Variable: VAR00001			
F	df1	df2	Sig.
2.819	1	37	.102

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.  
 a. Design: Intercept + VAR00002

Dari tabel 4 dapat dilihat bahwa nilai F sebesar 2,819 dan Sig. sebesar  $0,102 > 0,05$  maka data homogen. Dari hasil uji prasyarat dapat disimpulkan bahwa data berdistribusi normal dan homogen sehingga untuk uji perbedaan dua rata-rata dengan menggunakan uji T. Untuk uji perbedaan dua rata-rata untuk sampel bebas ini juga dilakukan dengan berbantuan *software* SPSS versi 20.

Hasil uji perbedaan dua rata-rata tersebut dapat dilihat pada tabel berikut ini:

Tabel 5. Uji perbedaan dua rata-rata untuk sampel bebas

	VAR00002	N	Mean	Std. Deviation	Std. Error Mean
VAR00001	1	19	.012142	.0029256	.0006712
	2	20	.011530	.0019892	.0004448

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
VAR00001	Equal variances assumed	2.819	.102	.768	37	.448	.0006121	.0007974	-.0010036	.0022279
	Equal variances not assumed			.760	31.522	.453	.0006121	.0008052	-.0010290	.0022532

Dari tabel 5 dapat dilihat bahwa nilai rata-rata koefisien reliabilitas *inter rater* dari guru matematika diperoleh sebesar 0,012142 sedangkan nilai rata-rata koefisien dari dosen pendidikan matematika sebesar 0,011530. Secara matematis nilai rata-rata koefisien reliabilitas *inter rater* yang diperoleh dari dosen pendidikan matematika lebih tinggi daripada yang diperoleh dari guru matematika SMP. Pada tabel Uji statistik pada tabel *independent sample test* diperoleh nilai F sebesar 2819, sig. sebesar 0,102 > 0,05 yang berarti varians kedua populasi sama atau homogen. Pada baris t-test for equality means diperoleh T sebesar 0,768 db = 37 dan sig.  $0,448/2 = 0,224 > 0,05$  yang berarti koefisien reliabilitas *inter rater* yang diperoleh dari guru matematika SMP secara signifikan tidak berbeda atau sama dengan koefisien reliabilitas *inter rater* yang diperoleh dari dosen pendidikan matematika.

## Pembahasan

Dari data yang diperoleh terdapat satu data yang *outlier* sehingga data tersebut dikeluarkan dari analisis. Tujuan pengeluaran data *outlier* tersebut untuk mengurangi kesalahan dan bias. Menurut Singh dan Jebalamar (2016: 789) bahwa mendeteksi *outlier* berarti mendeteksi data yang abnormal pada populasi yang terdefinisi untuk meningkatkan keamanan dan mengurangi kesalahan. Hal ini juga sejalan dengan pendapat Filzomer (2005:127) yang mengatakan bahwa data *outlier* dapat memungkinkan terjadinya bias dan dengan mengeluarkan data *outlier* tidak akan menghilangkan informasi yang diukur. Data yang dianalisis berjumlah 39 data yang terdiri dari 19 data yang berasal dari guru matematika dan 20 data yang berasal dari dosen pendidikan matematika.

Menurut Widhiarso (2010) melibatkan rater dalam pengembangan instrumen dapat meningkatkan kualitas alat ukur yang dikembangkan. Menurut Aisiah (2016), reliabilitas instrumen ditunjukkan oleh kekonsistenan penilaian antar *rater* untuk menilai suatu instrumen. Mengingat kompetensi penilai sangat mempengaruhi reliabilitas penilaian, maka penilai harus memiliki konsentrasi keahlian yang relevan dengan instrumen yang dinilai. Menurut pendapat Maesaroh (2016) yang mengatakan bahwa penilaian yang diberikan oleh rater itu sangat penting untuk meningkatkan kualitas instrumen sehingga perlu dipertimbangkan mengenai latar belakang *rater*.

Dari hasil yang diperoleh menunjukkan bahwa koefisien reliabilitas *inter rater* yang diperoleh dari guru matematika SMP secara signifikan tidak berbeda dengan koefisien reliabilitas *inter rater* yang diperoleh dari dosen pendidikan matematika, hal ini didukung oleh pendapat Aisyah (2016) dan Maesaroh (2016) bahwa latar belakang rater sangat penting untuk diperhatikan.

## 4. SIMPULAN DAN SARAN

### Simpulan

Tidak terdapat perbedaan yang signifikan antara koefisien reliabilitas *inter rater* soal tes matematika yang dinilai oleh guru matematika SMP dan dosen pendidikan matematika.

## Saran

Untuk kepentingan validasi suatu instrument yang akan menggunakan *rater* hendaknya diperhatikan bidang keahlian *rater* relevan dengan instrument yang akan dinilai.

## 5. DAFTAR PUSTAKA

- Aisiah. *Instrumen Penilaian Skripsi Jurusan Sejarah*. *jurnal.fkip.unila.ac.id*. <http://jurnal.fkip.unila.ac.id/index.php/jpp/article/view/5243/3301> (diakses 6 Maret 2016)
- Djanuarsih, Eri. 2012. *Validitas dan Reliabilitas Butir Soal*. *E-Jurnal Dinas Pendidikan Kota Surabaya*; Volume 1. No.1. ISSN : 2337-3253.
- Kadir. 2015. *Statistik Terapan konsep, Contoh dan Analisis Data dengan Program SPSS/Lisrel dalam Penelitian*. Jakarta : PT.Raja Grafindo Persada.
- Kartowagiran, Badrun. 2011. *Penulisan Dan Analisis Butir Soal*. *Makalah* disampaikan pada Pelatihan penulisan dan analisis butir bagi guru SMP Provinsi D.I. Yogyakarta pada tanggal 23, 26, dan 28 Juli 2011 di Pascasarjana UNY 2011. <http://staff.uny.ac.id> (diakses 29 Januari 2016)
- Kusaeri, K. 2014. *Acuan dan Tehnik Penilaian Proses dan hasil belajar dalam kurikulum 2013*. Yogyakarta: Ar-Ruzz Media.
- Maesaroh, Raden. *Teacher-Rater Language Academic Background Difference And Assessment Process Of Indonesian As Foreign Language (IFL) Speaking Performance (A Qualitative study conducted to three IFL teachers in one language center in Bandung)*. <http://File.Upi.Edu/> diakses 7 Maret 2016)
- Peter Filzmoser, 2005. "Identification of Multivariate Outliers: A Performance Study," *Australian Journal of Statistics*, Vol. 35, No. 2.
- Singh and Jebalamar, "Model-Based Outlier Detection System with Statistikal Preprocessing," *Journal of Modern Applied Statistikal Methods*, Vol. 15, No. 1, Mei 2016, h. 789.
- Widhiharso, wahyu. 2010. *Melibatkan Rater dalam Pengembangan Alat Ukur*. Fakultas Psikologi Universitas Gadjah Mada
- Widoyoko, S.Eko Putro, 2015. *Evaluasi Program pembelajaran*. Yogyakarta : Pustaka Pelajar.