

Estimasi Kesalahan Pengukuran dalam Penilaian Sidang Skripsi: Generalizability Theory Analysis

Islamiani Safitri^{1*}, Raden Rosnawati², Rizli Ansyari³, Rofia Abada⁴

¹Pendidikan Matematika, Universitas Labuhanbatu, Indonesia

²Penelitian dan Evaluasi Pendidikan, Universitas Negeri Yogyakarta, Indonesia

³Manajemen Pendidikan, Universitas Negeri Yogyakarta, Indonesia

⁴Abdelhafid Boussouf Mila University Center, Algeria

Email: islamiani.safitri@ulb.ac.id, rizliansyari.2022@student.uny.ac.id, r.abada@centre-univ-mila.dz

Abstrak

Tujuan dari penelitian ini adalah untuk mengestimasi kesalahan pengukuran dalam penilaian sidang skripsi mahasiswa Prodi Pendidikan Matematika, Universitas Labuhanbatu. Penelitian ini menggunakan pendekatan kuantitatif deskriptif untuk mengurai kesalahan pengukuran pada penilaian sidang skripsi mahasiswa. Populasi Populasi dalam penelitian ini adalah lembar penilaian sidang skripsi mahasiswa dan sampel yang digunakan adalah respon jawaban pada lembar penilaian sidang skripsi dari tiga dosen penguji yang menilai sebanyak 10 mahasiswa. Pengambilan sampel dilakukan dengan menggunakan teknik *proportionate stratified random sampling*. Setiap dosen penguji diberikan lembar penilaian dengan 8 item yang mengukur 8 aspek. Penilaian dari ketiga dosen penguji sidang skripsi akan dianalisis menggunakan *Generalizability Theory Two Facet* dengan desain $i \times r \times s$. Setelah dilakukan analisis, maka hasil dari penelitian ini yaitu varians terbesar berdasarkan G-study adalah pada aspek residual sebesar 75,1%, aspek interaksi student-rater sebesar 13,1%, dan aspek student sebesar 11,6%. Sedangkan ratio of varians-N yang paling besar adalah pada aspek student sebesar 5,12, lalu diikuti oleh aspek student-rater sebesar 0,96, aspek residual sebesar 0,23, dan aspek student-item sebesar 0,0045. Pada penelitian ini juga diperoleh *generalizability coefficient* awal sebesar 0.5096671, lalu dimodifikasi sehingga menghasilkan *dependability coefficient* sebesar 0.810694.

Kata Kunci: estimasi kesalahan pengukuran, generalizability theory, penilaian sidang skripsi

PENDAHULUAN

Setiap mahasiswa yang akan menyelesaikan pendidikannya, wajib mengerjakan tugas akhir. Pada jenjang sarjan (S1), tugas akhir biasa disebut dengan skripsi. Skripsi merupakan karya ilmiah berupa hasil penelitian yang dilakukan oleh mahasiswa sesuai dengan bidang keahliannya sebagai syarat lulus pendidikan sarjana (S1). Pada semester akhir, biasanya mahasiswa sudah disibukkan dengan melakukan penelitian yang akan dituangkan pada skripsi. Setelah mahasiswa menyelesaikan skripsinya, maka dilanjutkan dengan ujian skripsi yakni dengan melaksanakan sidang skripsi yang telah dijadwalkan oleh program studi masing-masing. Pada proses sidang skripsi, biasanya dihadirkan 1-3 dosen penguji untuk memberikan kritikan maupun saran sebagai masukan untuk skripsi agar hasilnya lebih baik. Dosen penguji terdiri dari dosen dengan bidang keilmuan yang sesuai dengan topik skripsi yang diangkat oleh mahasiswa.

Sama halnya dengan Program Studi Pendidikan Matematika di Universitas Labuhanbatu. Pelaksanaan sidang skripsi juga menghadirkan tiga dosen penguji yang terdiri atas 1 dosen pembimbing, 1 dosen penguji dengan bidang keilmuan matematika, dan 1 dosen

penguji bidang pendidikan. Setiap dosen penguji akan memberikan penilaian kepada mahasiswa sesuai dengan isi skripsi, penyajian skripsi pada saat presentase sidang, kemampuan mahasiswa dalam menjawab pertanyaan dari penguji, serta kemampuan mahasiswa dalam mempertahankan argumennya.

Namun, permasalahan muncul ketika ada kesenjangan antara nilai akhir dari sidang skripsi dengan kualitas skripsi maupun penyajian mahasiswa pada saat sidang skripsi. Nilai akhir sidang skripsi cenderung sama untuk setiap mahasiswa yakni dengan nilai A meskipun pada kenyataannya kualitas skripsi dan kemampuan mahasiswa dalam proses sidang skripsi sangat berbeda. Tentu saja hal ini memberikan ketidakadilan pada mahasiswa yang memang memiliki kualitas yang baik dalam menyusun skripsi maupun pada psroses sidang skripsinya. Sehingga muncul sebuah pertanyaan, apakah penilaian yang dilakukan oleh masing-masing dosen penguji yang dituangkan pada lembar penilaian sidang skripsi sudah mengungkapkan hasil yang sebenarnya? Benarkah disana tidak terjadi kesalahan dalam pengukuran? Karena pada kenyataannya, hasil penilaian antara satu dosen penguji dengan dosen penguji yang lain sering terdapat perbedaan atau rentang penilaian yang sangat jauh. Artinya, ada dosen penguji yang memberikan nilai sangat tinggi, namun di sisi lain ada pula dosen penguji yang memberikan nilai sangat rendah untuk mahasiswa yang sama. Hal ini menunjukkan bahwa ada kesalahan pengukuran (error) pada proses penilaian sidang skripsi. Namun belum diketahui apa yang menjadi sumber error pada penilaian tersebut. Sebab sampai saat ini belum pernah dilakukan uji coba maupun analisis untuk mengetahui error pada penilaian sidang skripsi. Oleh sebab itu, pada penelitian ini akan dilakukan analisis untuk mengestimasi kesalahan pengukuran pada penilaian sidang skripsi mahasiswa dengan pendekatan *Theory Generalizability two facet*.

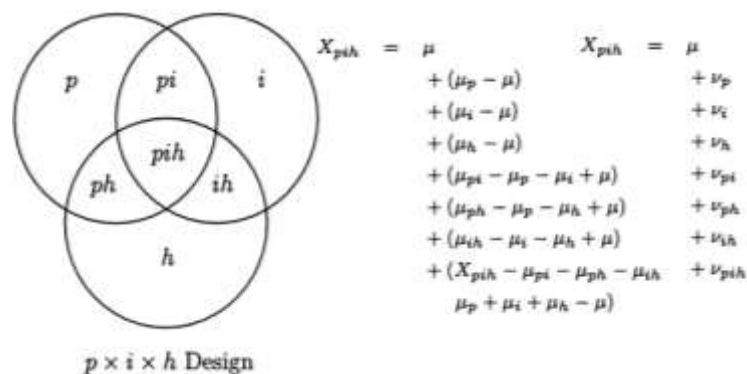
Pendekatan *Generalizability Theory* digunakan karena dalam metode *standard setting* ini digunakan panelis sebagai penentu *cut score* (Cruz, 2013; Huebner et al., 2019), dimana penggunaan panelis akan menimbulkan tingginya variabilitas nilai yang diberikan oleh panelis itu sendiri. Sehingga reliabilitas dari skor yang dihasilkan perlu diperhatikan. Penggunaan *generalizability theory* dikarenakan *G-theory* memberikan berbagai model yang digunakan dalam menyelidiki kesalahan. Analisis *G-theory* memiliki dua tahap yaitu generalisasi studi (G-study) dan decision studi (D-study) (Wendari & Hadi, 2018). G-study dilakukan untuk menentukan seberapa baik skor dapat digunakan dalam beberapa situasi dan melibatkan perkiraan komponen varians yang mung kin akan digunakan dalam studi-D. Studi D adalah penelitian yang dilakukan untuk menghitung koefisien reliabilitas dan SEs pengukuran dengan tujuan menentukan prosedur pengukuran yang paling efisien pada situasi tertentu. Pertimbangan studi D yang paling penting adalah spesifikasi dari generalisasi populasi dimana pembuat keputusan ingin menggeneralisasi skor dengan prosedur pengukuran tertentu.

METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif untuk mengestimasi kesalahan pengukuran pada penilaian sidang skripsi di Program Studi Pendidikan Matematika, Universitas Labuhanbatu. Sumber data pada penelitian ini merupakan mahasiswa tingkat akhir yang sedang melakukan ujian sidang skripsi di Program Studi Pendidikan Matematika, Universitas Labuhanbatu.

Populasi dalam penelitian ini adalah lembar penilaian sidang skripsi mahasiswa Program Studi Pendidikan Matematika di Universitas Labuhanbatu. Adapun sampel yang digunakan dalam penelitian ini merupakan respon jawaban pada lembar penilaian sidang skripsi dari tiga dosen penguji yang menilai sebanyak 10 mahasiswa. Pengambilan sampel dalam penelitian ini menggunakan teknik *proportionate stratified random sampling*. Setiap dosen penguji diberikan lembar penilaian dengan 8 item yang mengukur 8 aspek yaitu, 1) Kemampuan teknik penulisan skripsi, 2) Kelengkapan unsur skripsi, 3) Ruang lingkup dan kedalaman pembahasan, 4) Kecukupan dan kemutakhiran data/informasi dan metodologi, 5) Kelengkapan unsur dan kualitas sumber penelitian, 6) Kemampuan penyajian skripsi, 7) Kemampuan berargumentasi, dan 8) Kualitas PPT.

Hasil penilaian dari ketiga dosen penguji sidang skripsi akan dianalisis menggunakan Generalizability Theory Two Facet dengan desain $i \times r \times s$. Dimana i adalah item yaitu butir pertanyaan pada lembar penilaian, r adalah rater yaitu penguji pada sidang skripsi, dan s adalah student yaitu mahasiswa yang diuji pada sidang skripsi. Pada analisis ini akan menampilkan tujuh varians pengukuran berdasarkan interaksi dari ketiga variable tersebut. Adapun varians yang dimunculkan adalah varians item, varians rater, varians student, varians student-rater, varians item-student, varians rater-item, dan kesalahan residual. Dari analisis ini akan diketahui variable mana yang memiliki varians paling besar memberikan *error* pada penilaian sidang skripsi. Berikut adalah diagram venn *Generalizability Theory Two Facet* dengan desain $i \times r \times s$.



Gambar 1. Diagram venn Generalizability Theory Two Facet dengan desain $i \times r \times s$
 (Brennan, 2001)

HASIL DAN PEMBAHASAN

Pada penelitian ini terdapat 10 mahasiswa melaksanakan sidang skripsi di Program Studi Pendidikan Matematika, Universitas Labuhanbatu. Masing-masing mahasiswa tersebut diuji oleh tiga dosen penguji dengan lembar penilaian yang sama. Lembar penilaian tersebut terdiri dari delapan item yang mengukur: 1) kemampuan teknik penulisan skripsi, 2) kelengkapan unsur skripsi, 3) ruang lingkup dan kedalaman pembahasan, 4) kecukupan dan kemutakhiran data/informasi dan metodologi, 5) kelengkapan unsur dan kualitas sumber penelitian, 6) kemampuan penyajian skripsi, 7) kemampuan berargumentasi, dan 8) kualitas PPT. Selanjutnya, data-data tersebut dianalisis menggunakan *Generalizability Theory Two Facet*.

dengan desain $i \times r \times s$. Berikut adalah hasil analisis *Generalizability Theory Two Facet* melalui Program R.

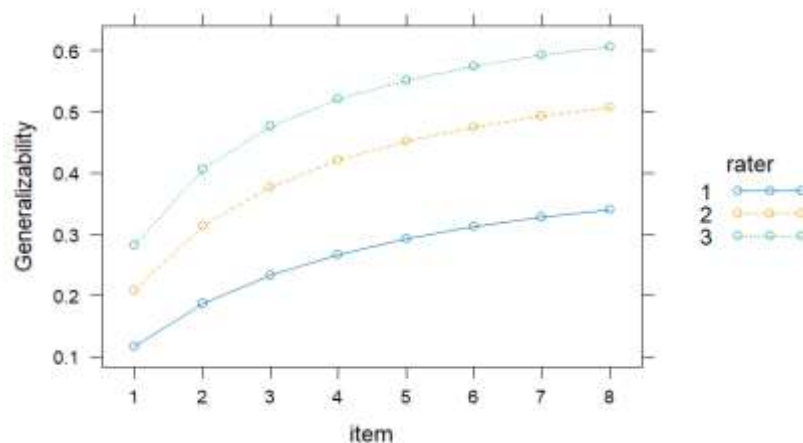
Table 1. Varians dari G-Study ($i \times r \times s$)

Source	Est.Variance	Percent.Variance
1 student:item	0.1080	0.2%
2 student:rater	5.7887	13.1%
3 item:rater	0.0000	0%
4 student	5.1397	11.6%
5 item	0.0000	0%
6 rater	0.0000	0%
7 Residual	33.2486	75.1%

Data pada Tabel 1 memberikan informasi bahwa terdapat 4 dari 7 aspek yang memberikan varians pada penilaian sidang skripsi yaitu, aspek interaksi student-item sebesar 0,2%, aspek interaksi student-rater sebesar 13,1%, aspek student sebesar 11,6%, dan aspek residual sebanyak 75,1%. Dari data tersebut dapat diketahui bahwa varians terbesar ditunjukkan oleh aspek residual lalu diikuti oleh aspek interaksi antara student-rater, aspek student, dan interaksi antara student-item.

Varians pada interaksi antara student-rater menunjukkan bahwa dosen penguji (rater) memberikan penilaian kepada beberapa mahasiswa lebih tinggi dibandingkan dengan mahasiswa lainnya. Artinya, terdapat dosen penguji yang tidak konsisten dalam memberikan penilaian kepada mahasiswa. Biasanya hal ini terjadi pada dosen penguji yang statusnya juga sebagai dosen pembimbing mahasiswa yang sedang melaksanakan sidang skripsi tersebut. Jadi, dosen pembimbing yang sekaligus berperan sebagai dosen penguji akan memberikan penilaian yang cenderung lebih tinggi dibandingkan dengan dosen penguji yang lain meskipun nilai yang diberikan tidak sesuai dengan kemampuan yang dimiliki oleh mahasiswa tersebut. Hal ini terjadi dengan alasan bahwa dosen pembimbing hendak membantu ataupun mendongkrak nilai mahasiswa bimbingannya. Jadi, jika beberapa dosen penguji memberikan nilai yang rendah kepada mahasiswa saat melaksanakan sidang skripsi, maka dosen pembimbingnya akan memberikan nilai yang sangat tinggi untuk mendongkrak nilai mahasiswa tersebut sehingga bisa mendapatkan nilai A. Seharusnya hal ini tidak boleh terjadi karena dapat menyebabkan penilaian sidang skripsi yang tidak objektif (Azwar, 2012; Istiyono, 2020; Mardapi, 2016; Retnawati, 2015).

Selanjutnya, varians terbesar pada Tabel 1 ditunjukkan oleh aspek residual yaitu sebesar 75,1%. Varians pada residual menunjukkan bahwa ada hal lain yang mempengaruhi dalam penilaian sidang skripsi mahasiswa yang tidak diperhitungkan dalam analisis penelitian ini. Jadi, meskipun kesalahan pengukuran (*error*) pada interaksi antara student-rater cukup tinggi, sebenarnya masih ada pengaruh luar atau hal lain yang juga mempengaruhi varians maupun *error* pada penilaian sidang skripsi di Program Studi Pendidikan Matematika Universitas Labuhanbatu. Dari Tabel 1 juga diketahui bahwa aspek item pada lembar penilaian sidang skripsi tidak memberikan *error* pada penilaian tersebut. Justru, varians student dalam hal ini adalah mahasiswa yang memberikan *error* pada pengukuran yakni sebesar 11%. Berikut adalah plot G-Study berdasarkan analisis *Generalizability Theory Two Facet*.



Gambar 2. Plot G-Study dari analisis Generalizability Theory Two Facet ($i \times r \times s$)

Meskipun angka varians pada aspek student cukup kecil (11%), namun hal ini dapat dikaitkan dengan varians pada aspek interaksi student-rater, yaitu keterkaitan khusus antara mahasiswa dengan dosen pembimbingnya. Oleh sebab itu, yang perlu diperbaiki pada penilaian sidang skripsi adalah konsistensi dosen penguji (*rater*) dalam memberikan penilaian kepada mahasiswa. Upaya sosialisasi terkait pentingnya penilaian secara objektif pada sidang skripsi perlu dilakukan agar dapat menghilangkan penilaian-penilaian yang mempertimbangkan faktor kedekatan antara mahasiswa dengan dosen pembimbingnya, juga agar hasil penilaian yang didapatkan sudah mengukur secara tepat aspek yang diukur dan tidak memberikan error yang begitu besar dalam penilaian. Selain itu, pada analisis ini juga ditemukan bahwa koefisien G-Study atau *generalizability coefficient* pada penelitian ini sebesar 0.5096671. Angka ini menunjukkan reliabilitas yang rendah terhadap penilaian sidang skripsi Prodi Pendidikan Matematika, Universitas Labuhanbatu.

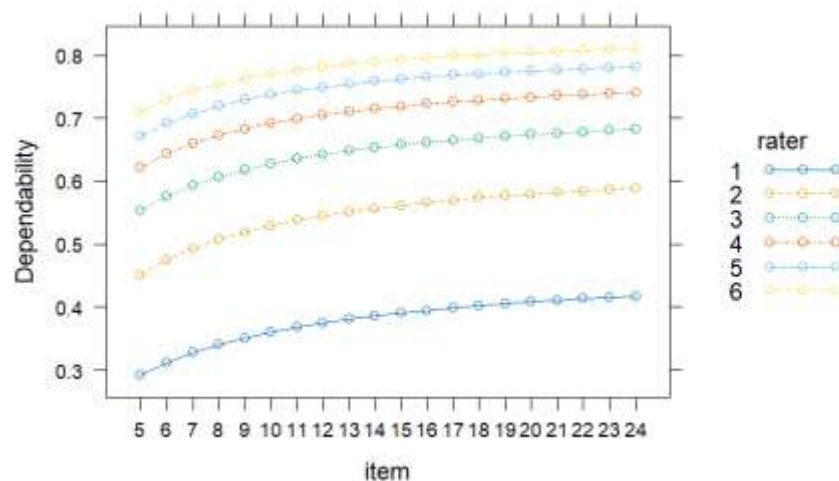
Setelah diketahui varians G-Study dan koefisien generalizability ini, maka dilanjutkan dengan melakukan analisis untuk menentukan D-Study. Jika estimasi G-Study dapat digunakan untuk merancang prosedur pengukuran yang efisien, maka D-Study digunakan untuk memberikan informasi dan membuat keputusan substantif tentang objek pengukuran (Hapsan & Rosnawati, 2022; Retnawati, 2015). Oleh karena itu, pada penelitian ini D-Study menekankan pada simulasi dalam memperbaiki *error* penilaian sidang skripsi berdasarkan varians terbesar pada G-Study.

Simulasi pada analisis D-study ini dilakukan dengan cara memodifikasi variable terkait, dalam hal ini yang dimodifikasi adalah jumlah rater dan item yang ditingkatkan. Pada analisis D-study, rater diperbanyak jumlahnya menjadi 6 orang, dan item diperbanyak menjadi 24 item. Berikut adalah hasil simulasi dari analisis D-Study.

Tabel 2. Varians D-Study

Source	Est.Variance	N	Ratio of Var:N
1 student:item	0.1080	24	0.0045000
2 student:rater	5.7887	6	0.9647833
3 item:rater	0.0000	144	0.0000000
4 student	5.1397	240	5.1397000
5 item	0.0000	24	0.0000000
6 rater	0.0000	6	0.0000000
7 Residual	33.2486	144	0.2308931

Berdasarkan data pada Tabel 2 diperoleh rasio dari varians-N terbesar berada pada aspek student sebesar 5,14, lalu diikuti oleh aspek student-rater sebesar 0,96, aspek residual sebesar 0,23 dan aspek student-item sebesar 0,0045. Dari varians D-Study ini dapat diketahui bahwa aspek-aspek yang memiliki varians pada G-Study yang besar, belum tentu memperoleh D-Study yang besar. Pada aspek student yang memiliki varians hanya 11% pada G-study justru memiliki D-study yang paling tinggi diantara aspek yang lain (Planinic et al., 2019; Retnawati, 2016; Ridwan et al., 2023). Selanjutnya, plot D-study dapat dilihat melalui gambar berikut.



Gambar 3. Plot D-Study dari analisis Generalizability Theory Two Facet ($I \times r \times s$)

Dari plot tersebut diketahui bahwa terjadi pengurangan *error* ketika rater dan jumlah item ditambah yakni rater menjadi 6 orang dan item mejadi 24 butir. Selain itu, pada D-study juga menghasilkan *dependability coefficient* sebesar 0.810694. Artinya, angka ini sudah cukup baik dibandingkan dengan koefisien generalizability awal yang hanya 0,51. Oleh karena itu, hal ini ini dapat disimpulkan bahwa estimasi kesalahan pengukuran pada siding skripsi dapat dihindari jika rater dan jumlah item pada lembar penilaian ditambah.

KESIMPULAN

Varians terbesar berdasarkan G-study adalah pada aspek residual sebesar 75,1%, aspek interaksi student-rater sebesar 13,1%, dan aspek student sebesar 11,6%. Sedangkan ratio of varians-N yang paling besar adalah pada aspek student sebesar 5,12, lalu diikuti oleh aspek student-rater sebesar 0,96, aspek residual sebesar 0,23, dan aspek student-item sebesar 0,0045. Pada penelitian ini juga diperoleh *generalizability coefficient* awal sebesar 0.5096671, lalu dimodifikasi sehingga menghasilkan *dependability coefficient* sebesar 0.810694.

DAFTAR PUSTAKA

- Azwar, S. (2012). *Reliabilitas dan Validitas* (4th ed.). Pustaka Belajar.
- Brennan, R. L. (2001). *Statistics for Social Science and Public Policy: Generalizability Theory*. Springer-Verlag.
- Cruz, A. P. S. (2013). Studi Generalizabilitas Tes Tipe Dua Facet Dengan Menggunakan Analisis Varian Tiga Jalur. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
- Hapsan, A., & Rosnawati, R. (2022). Koefisien Realibilitas dengan Teori Generalizabilitas (*G - Theory*) *t wo-Facet i : h : p design*. 1, 1–15.
- Huebner, A., Huebner, A., Lucht, M., & Dame, N. (2019). *Generalizability Theory in R* *Generalizability Theory in R*. 24(November).
- Istiyono, E. (2020). *Pengembangan Instrumen Penilaian dan Analisis Hasil Belajar Fisika: Dengan Teori Tes Klasik dan Modern*. UNY Press.
- Mardapi, D. (2016). *Pengukuran Penilaian dan Evaluasi Pendidikan*. Nuha Litera.
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research : Why measurement matters. *Physical Review Physics Education Research*, 15(2), 20111. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Retnawati, H. (2015). *Validitas dan Reliabilitas Karakteristik Butir (Panduan untuk peneliti, mahasiswa, dan psikometri)*. Parama Publishing.
- Retnawati, H. (2016). Proving Content Validity of Self-Regulated Learning Scale (The Comparison of Aiken Index and Expanded Gregory Index). *Research and Evaluation in Education*, 2(2), 155–164. <https://doi.org/10.21831/reid.v2i2.11029>
- Ridwan, M. R., Hadi, S., Jailani, J., & Retnawati, H. (2023). *The instrument development to measure the verbal ability of prospective high school students*. 12(1). <https://doi.org/10.11591/ijere.v12i1.22736>
- Wendari, W., & Hadi, S. (2018). Estimasi kesalahan pengukuran standard setting dalam penilaian kompetensi matematika tingkat SMP di Kabupaten Sumbawa. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 22(1), 24–34. <https://doi.org/10.21831/pep.v22i1.16492>