

# Case Study: Assessing NLX Coverage of QCEW School Employers in Minnesota

Amanda Rohrer  
Analyst Resource Center (ARC)  
State of Minnesota

The NLX Research Hub data contains a massive amount of data from employers. They essentially self-select by using a job posting service and can move in and out of the universe. To meaningfully compare it to statistical sources of data, it's necessary to have some sense of scope relative to a known universe of employers. Unfortunately, employers advertise jobs under other names and other hierarchies of ownership than they use to report their data to state and federal agencies. Job postings are also advertisements, not jobs. They sometimes run unfilled for several months and can be distributed widely (resulting in geographic duplication). They're compiled from multiple sources – employers paying for a service and state job banks – and as a result coverage may vary between states. Making the comparison to established data programs requires custom matching of entities between the two sources.

To see what kind of coverage the NLX data contains and the amount and type of work it would take to create a crosswalk between companies advertising jobs on NLX and the QCEW universe of liable employers, we used Schools (NAICS 6111), both public and private, as a test case for matching the two data sources.

## Definitions:

*Jobhistory* – This is a file format provided by the Research Hub that includes advertisements less descriptions organized by the day they were advertised (*file\_date*). To manage data size, only one day a week was brought in (Mondays). The time series goes back to mid-2021. The API provides full text of descriptions but they must be selected based on when jobs are first created, which can be long ago for evergreen jobs and listings that are reused rather than created new. For this analysis, the *jobhistory* files were the source of NLX data.

*Application\_company* – the field name used to describe an employer name in the NLX source data. FEIN has incomplete coverage and there is no other employer identifier so this is treated as the primary identifier for employers in the NLX.

*Trade\_name* and *Legal\_name* – the fields QCEW uses for names of a business. One has to be populated but either can be blank.

*QCEW Employer* – Distinct UI Accounts (singles and parent units) with quarterly average employment of greater than 10 and assigned to NAICS 6111 for the reference quarter (2023 Q2). The very small employers were excluded for two reasons. Their hiring is so infrequent it's difficult to determine if they're absent because they don't use the NLX service or because they have no openings. They also often represent tutoring services or a remote worker or a not-yet-open charter school rather than a true school.

*Match* – a confirmed relationship between an *application\_company* and a QCEW Employer

## Considerations:

Schools were chosen as the test case because they minimize the impact of known characteristics of the NLX data. Schools are local, don't cross state boundaries, and are very physical-location based. This makes it likely that the location in which their job postings are advertised is similar to where they are located and limits the amount of duplication that may come from a widely advertised position. Schools are also present in all parts of the state, allowing more meaningful geographic comparison of coverage.

The industry is relatively small, making the amount of required matching attainable even if it needs to be somewhat manual. The industry has distinct economic and seasonal trends – patterns that can be compared in multiple data sources. It also has occupations (teachers – SOCs beginning 25-202X or 25-203X) that are almost unique to the industry, which is useful for identifying probable schools in the NLX data where NAICS code is very poorly reported.

Schools are also mostly geographically clustered, which made using the parent UI Account reasonable in a match. For other industries like Retail Trade matching would have to be on both city and application\_company.

Even the largest schools don't advertise jobs every day of the year. When selecting the application\_companies from NLX to reference, it was necessary to use the entire span of the available jobhistory files in order to get meaningful matches. For a point-in-time comparison, the weekly average for the reference was used and for time-series comparison, the total for the day closest to the end of the month was used.

Relationships are many-to-many. Some districts will group their schools by elementary and secondary even though they're a single district with several units in the QCEW universe. Sometimes there are minor differences in how a name shows up in application\_company (e.g. X Public Schools vs X School District). This affects how the data is aggregated and needs to be taken into consideration to avoid potential duplication.

The table below is the structure of the crosswalk table that was created to store "matches". Blue fields came from the NLX data. The green fields come from QCEW microdata and are obscured in this report because they're not public. The additional field of matchtype was created to track how the match was made. Any employer can appear multiple times as an application\_company or as a uiaccount. All associated reporting units were included. Even though reporting units were not used for this analysis, the structure is intended to be useful for future efforts, as well.

application_company	fein	matchtype	uiaccount	reptunit	trade_name	legal_name
Ethan Allen Global Inc.	XXXXXXXXXX	fein	XXXXXXXX	XXXXX	HIDDEN	HIDDEN
Family Dollar	XXXXXXXXXX	trade name match	XXXXXXXX	XXXXX	HIDDEN	HIDDEN
Zwicker & Associates, P.C.	XXXXXXXXXX	manual	XXXXXXXX	XXXXX	HIDDEN	HIDDEN
7-Eleven	XXXXXXXXXX	legal name match	XXXXXXXX	XXXXX	HIDDEN	HIDDEN
ZWILLING J.A. Henckels	XXXXXXXXXX	trade name match	XXXXXXXX	XXXXX	HIDDEN	HIDDEN
The Capital Grille	XXXXXXXXXX	legal name match	XXXXXXXX	XXXXX	HIDDEN	HIDDEN
Universal Forest Products	XXXXXXXXXX	trade name contains	XXXXXXXX	XXXXX	HIDDEN	HIDDEN
Universal Forest Products Minneota, LLC	XXXXXXXXXX	trade name match	XXXXXXXX	XXXXX	HIDDEN	HIDDEN

## Method:

To match QCEW employers with application\_companies, we began with the most automatic and gradually got to more manual methods. These first steps were applied to all companies and industries. Some companies include their FEIN in the NLX data. Because that's also available in the QCEW universe, those were matched first and considered most reliable. As a second step, a direct match of the application\_company to either the trade or legal name used in QCEW was used. The next step was to identify records where the entire application\_company was contained in the trade or legal name. While this was useful for removing variations like LLC and Inc., some company names are reported as short acronyms and can result in a false positive. Those were reviewed and some company names were manually excluded from that match. Then some known patterns were matched – for example, county government often uses a name like “X County” in NLX but is “County of X” in Minnesota’s QCEW data. That pattern is easy to identify and match once known.

Next, the test employers were identified. All Minnesota employers that were identified in the reference quarter as belonging to the NAICS code 6111, meeting the employment threshold, and having no existing match were selected and manually assigned an NLX match if one could be found. Many districts have obvious keywords and searching was more time-consuming than onerous. Locations were sometimes helpful in distinguishing similar names. Ranking in reverse order of employment helped the process.

Eventually the frequency of found matches drops off. At that point, application\_companies that had hired teachers as represented by any previous job posting with an associated ONET code with the 25-202X.XX or 25-203X.XX pattern and that had not been matched were identified. Those were matched from the NLX application\_company to the QCEW employer. Ultimately, fewer than 20 application\_companies matching those criteria remained and were either summer camps, educational services, or parochial schools that had opted out of unemployment insurance in Minnesota and therefore were correctly not in the QCEW universe.

Once all were matched, the data was reviewed for outliers. One useful technique for that was to chart weekly postings for each application company in a time series. This identified a few matches that needed to be reviewed where, for example, all the employment for a county was being matched to the school.

## Results:

Overall 22.8 percent of QCEW schools could be found in the NLX data, representing 54.2 percent of employment. The schools that were using NLX tended to be larger than those that did not, with 54.0 percent of those with more than 200 employees being found and only 4.4 percent of those with fewer than 50 employees being found. When compared to postings limited to the reference quarter instead of the entire span of data, fewer schools were represented, only about a third. It's not possible to determine if the absence of an employer in a given time frame is because they stopped using the service or because they had no openings at that time, but the stronger representation of large schools in the matched data is more pronounced in the shorter time series.

The relative measure of NLX-advertised job postings to QCEW employment was also higher for larger employers – whether that reflects how the schools are using the NLX or other characteristics of hiring in small establishments would require further investigation.

Geographically, the rate of matching was much more consistent, ranging from 17.0 percent in Northwest Minnesota to 32.6 percent in Northeast Minnesota. The Twin Cities, where most of the state's population

resides, was in the middle at 24.3 percent. The availability of matches throughout all regions is particularly encouraging for future use cases.

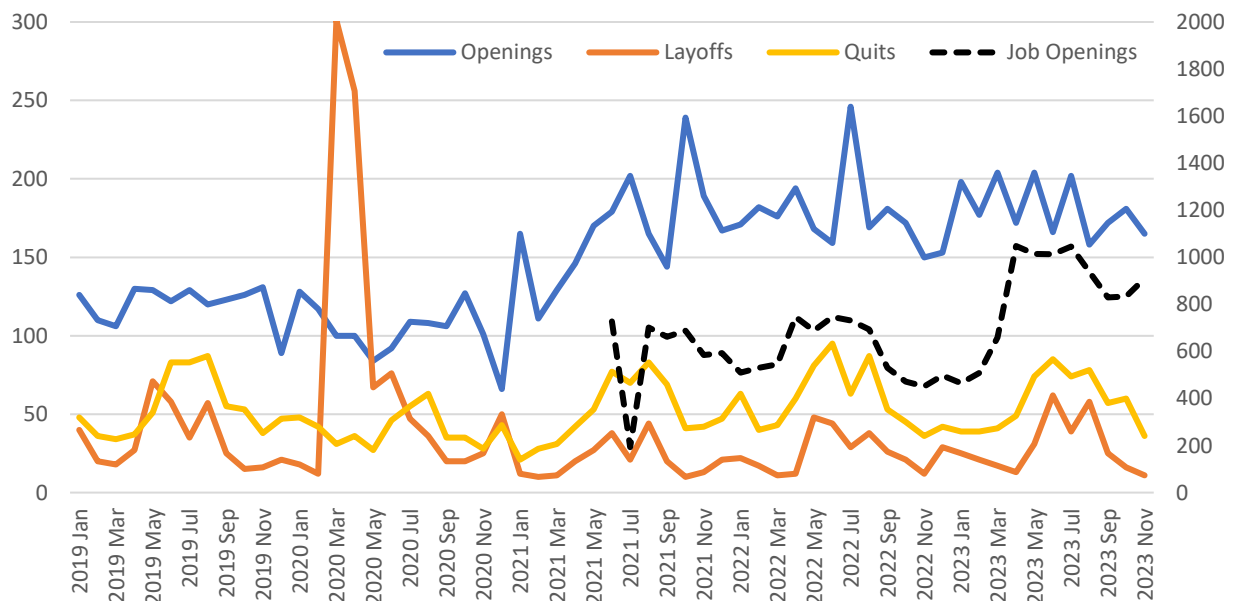
### 2023 Q2 Comparison of QCEW Employment to Matched NLX Job Posting Data NAICS 6111

	Total	Matched	Matched in Same Quarter	Pct Matched	Pct Matched in Same Quarter	Employment	Matched Employment	Pct Matched	Weekly Avg Q2 2023 openings	Rate of Opening
<b>Employer Size</b>	640	146	55	22.8%	8.6%	163,723	88,750	54.2%	932	0.6%
<50	183	8	1	4.4%	0.5%	5,438	266	4.9%	7	0.1%
50-100	136	24	8	17.6%	5.9%	10,125	1,904	18.8%	29	0.3%
100-200	160	27	12	16.9%	7.5%	22,538	3,820	16.9%	55	0.2%
200+	161	87	34	54.0%	21.1%	125,622	82,759	65.9%	841	0.7%
<b>Region</b>										
Central	79	21	5	26.6%	6.3%	21,915	9,763	44.6%	140	0.6%
NE	43	14	7	32.6%	16.3%	8,370	5,075	60.6%	30	0.4%
NW	112	19	9	17.0%	8.0%	17,776	4,994	28.1%	63	0.4%
SE	72	14	3	19.4%	4.2%	15,118	6,226	41.2%	1	0.0%
SW	81	20	9	24.7%	11.1%	13,131	4,701	35.8%	7	0.1%
Twin Cities	251	61	22	24.3%	8.8%	87,363	57,990	66.4%	689	0.8%

As a time series, the total openings matched to a school were compared to the national JOLTS data for the Educational Services industry. That’s a larger industry than just schools, including colleges and trade schools, but in Minnesota 6111 makes up the largest share of that total.

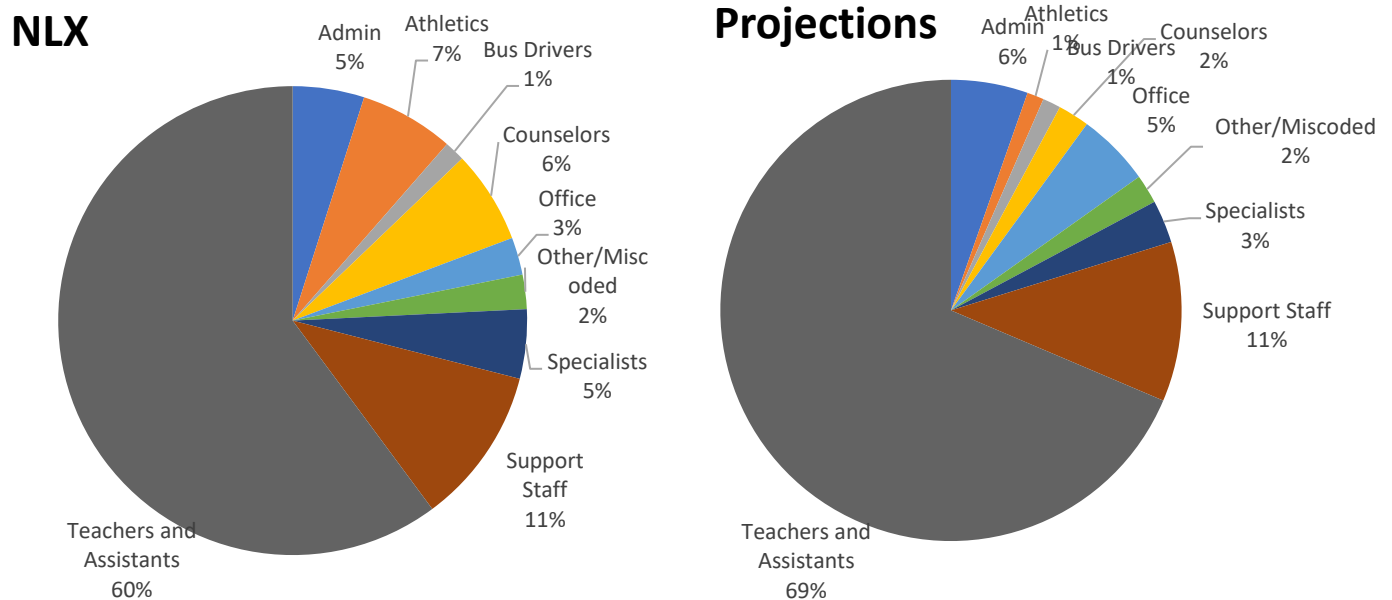
The NLX openings series has a stronger upward trend than the national data. That could be thanks to the absence of colleges or a product of local trends. It may also change if the total was weighted for size and region. The NLX openings show a similar seasonal trend to Quits and Layoffs in the JOLTS data.

Matched Schools Openings vs JOLTS Educational Services National Trend



Finally, the proportion of openings by NLX-assigned occupation (excluding null codes) was compared to the distribution of jobs by occupation in Projections matrix data for NAICS 6111 in Minnesota. The two sources have some significant differences conceptually – projections is existing jobs, while NLX is advertised openings. Different types of jobs have different turnover rates, meaning they could be hired more frequently without having more actual jobs performing a function. Some may remain open for longer than others, which would mean that, when we look at all openings through time the longer-running openings would be overstated relative their proportion in Projections. Some jobs may be advertised differently, such as through unions or professional organizations or employer services. Regardless, it’s useful to know how the two sources line up to understand the scope of the data. In the charts below, NLX shows somewhat fewer Teachers and Assistants and Administrators relative to the Projections proportion. Counselors and Specialists (such as Speech Pathologists), which may be hard to fill, are somewhat over-represented. While the distribution is not perfect, it’s reassuringly close.

Aggregate NLX Job Openings by Job Type for Schools (6111)



**Conclusions:**

Coverage is an essential question that needs to be explored to meaningfully use the NLX data. While it’s possible to line it up with established data sources, the opt-in nature of the data means that there will be variations based on industry characteristics and likely between states, as well. Employers can also enter and leave the pool of NLX employers at any time – review needs to be ongoing.

This analysis of schools is promising – there’s reasonable coverage with good geographic distribution, occupations being advertised are consistent with the industry and seasonal trends are present. There weren’t extra application\_companies on the NLX side that couldn’t be matched to a QCEW employer. Even so, schools were specifically chosen because they’re different than other industries, being more consistently spread across the state, having fewer remote workers, and no multi-state locations. Similar comparisons made in other industries and other states are necessary.